HANDBOOKS IN ECONOMICS

Anthony B. Atkinson
François Bourguignon

Volume 2
2015

Handbook of Income Distribution

NORTH-HOLLAND

Handbook of
INCOME DISTRIBUTION

Volume 2

First edition

Edited by

ANTHONY B. ATKINSON
Nuffield College, Oxford, UK

FRANÇOIS BOURGUIGNON
Paris School of Economics, Paris, France

2015

ELSEVIER

Amsterdam • Boston • Heidelberg • London • New York • Oxford
Paris • San Diego • San Francisco • Singapore • Sydney • Tokyo
North-Holland is an imprint of Elsevier

N·H

Table of Contents

CONTRIBUTORS

Rolf Aaberge
Statistics Norway, Research Department and University of Oslo, Department of Economics, ESOP, Oslo, Norway

Daron Acemoglu
Massachusetts Institute of Technology, Cambridge, MA, USA

Facundo Alvaredo
EMod/OMI-Oxford University, Paris School of Economics, and CONICET, Oxford, UK

Sudhir Anand
Department of Economics, University of Oxford, Oxford, UK; and Department of Global Health and Population, Harvard School of Public Health, Boston, MA, USA

Andrea Brandolini
Bank of Italy, Directorate General for Economics, Statistics, and Research, Rome, Italy

Daniele Checchi
Department of Economics, University of Milan, Milano, Italy and IZA, Bonn, Germany

Pierre-André Chiappori
Department of Economics, Columbia University, NY, USA

Andrew E. Clark
Paris School of Economics—CNRS, Paris, France

Frank A. Cowell
STICERD, London School of Economics, London, UK

Conchita D'Ambrosio
INSIDE, Université du Luxembourg, Walferdange, Luxembourg

Koen Decancq
Centre for Social Policy Herman Deleeck, University of Antwerp, Antwerp; Department of Economics, Katholieke Universiteit Leuven, Leuven, and CORE, Université catholique de Louvain, Louvain-la-Neuve, Belgium

Jean-Yves Duclos
IZA, Département d'économique and CIRPÉE, Université Laval, Quebec, QC, Canada

Francesco Figari
University of Insubria, Varese, Italy, and Institute for Social and Economic Research (ISER), University of Essex, Colchester, UK

Emmanuel Flachaire
Aix-Marseille Université, AMSE & IUF, Marseille, France

Marc Fleurbaey
Princeton University, Princeton, NJ, USA

Michael F. Förster
OECD, Paris, France

Leonardo Gasparini
CEDLAS-FCE-Universidad Nacional de La Plata, and CONICET, La Plata, Argentina

István György Tóth
Tárki Social Research Institute, Budapest, Hungary

Markus Jäntti
Swedish Institute for Social Research (SOFI), Stockholm University, Stockholm, Sweden

Stephen P. Jenkins
Department of Social Policy, London School of Economics, London; ISER, University of Essex, Colchester, UK, and IZA, Bonn, Bonn, Germany

Ravi Kanbur
Cornell University, Ithaca, NY, USA

Ive Marx
Herman Deleeck Centre for Social Policy, University of Antwerp, Antwerp, and Institute for the Study of Labor, IZA, Bonn, Germany

Costas Meghir
Department of Economics, Yale University, New Haven, CT, USA

Dominique Meurs
University Paris Ouest Nanterre, EconomiX and Ined, France

Salvatore Morelli
CSEF, University of Naples, Federico II, Naples, Italy, and INET Oxford, University of Oxford, Oxford, UK

Suresh Naidu
Columbia University, NY, USA

Brian Nolan
Department of Social Policy and Intervention and Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, UK

Owen O'Donnell
Erasmus School of Economics & Tinbergen Institute, Erasmus University Rotterdam, Rotterdam, The Netherlands, and School of Economics and Area Studies, University of Macedonia, Thessaloniki, Greece

Javier Olivera
Institute for Research on Socio-Economic Inequality, RU INSIDE, University of Luxembourg, Luxembourg, Luxembourg

Alari Paulus
Institute for Social and Economic Research (ISER), University of Essex, Colchester, UK

Thomas Piketty
Paris School of Economics, Paris, France

Sophie Ponthieux
Institut national de la statistique et des études économiques, France

Vincenzo Quadrini
University of Southern California, CEPR, Los Angeles, CA, USA

Martin Ravallion
Department of Economics, Georgetown University, Washington, DC, USA

Pascual Restrepo
Massachusetts Institute of Technology, Cambridge, MA, USA

José-Víctor Ríos-Rull
University of Minnesota, Federal Reserve Bank of Minneapolis, CAERP, CEPR, NBER, Minneapolis, MN, USA

James A. Robinson
Harvard University, Cambridge, MA, USA

John E. Roemer
Yale University, New Haven, CT, USA

Jesper Roine
Stockholm Institute of Transition Economics, Stockholm School of Economics, Stockholm, Sweden

Wiemer Salverda
Amsterdam Institute for Advanced Labour Studies (AIAS), and Amsterdam Centre for Inequality Studies (AMCIS), University of Amsterdam, Amsterdam, The Netherlands

Agnar Sandmo
Department of Economics, Norwegian School of Economics (NHH), Bergen, Norway

Erik Schokkaert
Department of Economics, Katholieke Universiteit Leuven, Leuven, and CORE, Université catholique de Louvain, Louvain-la-Neuve, Belgium

Paul Segal
King's International Development Institute, King's College London, London, UK

Timothy Smeeding
University of Wisconsin, Madison, WI, USA

Holly Sutherland
Institute for Social and Economic Research (ISER), University of Essex, Colchester, UK

André-Marie Taptué
Département d'économique and CIRPÉE, Université Laval, Quebec, QC, Canada

Jeffrey Thompson
Federal Reserve Board of Governors, Washington, DC, USA

Alain Trannoy
Aix-Marseille University, EHESS and CNRS, France

Eddy Van Doorslaer
Erasmus School of Economics, Tinbergen Institute & Institute for Health Policy and Management, Erasmus University Rotterdam, Rotterdam, The Netherlands

Tom Van Ourti
Erasmus School of Economics & Tinbergen Institute, Erasmus University Rotterdam, Rotterdam, The Netherlands

Daniel Waldenström
Department of Economics, Uppsala University, Uppsala, Sweden

Gabriel Zucman
London School of Economics, London, UK

INTRODUCTION: INCOME DISTRIBUTION TODAY

Anthony B. Atkinson*, François Bourguignon†
*Nuffield College, Oxford, UK
†Paris School of Economics, Paris, France

1. SETTING THE SCENE

When the first volume of the *Handbook of Income Distribution* was published in 2000, the subject of income inequality was not in the mainstream of economic debate—despite the long history of engagement with this issue by earlier leading economists—see Chapter 1 by Agnar Sandmo. Fifteen years later, inequality has become very much centre stage. Rising income inequality has attracted the attention of the U.S. President, of international bodies such as the IMF and the OECD, and of participants in the Davos meeting.

This volume of the Handbook aims to cover the advances made in the past 15 years in our understanding of the extent, causes, and consequences of inequality. In this respect, the second volume should be seen as complementing, not supplanting, the first volume. We have encouraged authors to concentrate on the developments that have taken place since 2000, and the chapters should be read in conjunction with those in volume I. In this Introduction, we give a flavor of the issues discussed and some personal reflections on the state of the subject.

In our Introduction to volume I of the Handbook, we said that "income distribution may be considered the normative economic issue 'par excellence'" (Atkinson and Bourguignon, 2000, p. 41). People are concerned about economic inequality because they feel it to be socially unjust or unfair. It violates principles of social justice. The nature of these principles is of course much debated and there is disagreement about what constitutes an unacceptable level of inequality. People focus on different dimensions. But the concern is with inequality intrinsically. At the same time, there is a second set of—instrumental—concerns with the consequences of inequality. The societal consequences were highlighted by Joseph Stiglitz when he entitled his 2012 book, *The price of inequality*, where he says "the impact of inequality on societies is now increasingly well understood—higher crime, health problems, and mental illness, lower educational achievements, social cohesion and life expectancy" (inside cover). The social, political and cultural impacts of inequality have been the subject of the GINI (Growing

Inequalities' Impacts) research project (Salverda et al., 2014; Nolan et al., 2014). In the second volume of the Handbook, some of these consequences feature, notably those regarding health in Chapter 17 by Owen O'Donnell, Eddy Van Doorslaer, and Tom Van Ourti. But the wider societal impact of inequality is not the principal focus of the chapters that follow. This still leaves much to be discussed. Inequality is of instrumental importance within the field of economics itself. As we said in volume I, "income distribution assists our understanding of various fields of economics" (2000, p. 4). Now, as then, we believe that the study of economic inequality should be at the heart of economic analysis.

Upon reflecting the issues covered in this volume, this Introduction considers successively: (a) the concepts and approaches to economic inequality measurement, or the various facets of inequality (Section 2); (b) the care needed with data on inequality (Section 3); (c) the explanations of changes in various dimensions of economic inequality, most notably the distribution of income, earnings and wealth and the links with macroeconomics (Section 4); and (d) the policies available to influence those changes or to correct those distributions (Section 5).

We plunge straight into the subject matter with Figure 1, which depicts the evolution of economic inequality in the United States over the past century. The data are taken from the Chartbook of Economic Inequality (Atkinson and Morelli, 2014),[1] but are presented in four panels to highlight different dimensions of the distribution of income.[2] The pictures provide a good basis for describing what is covered in this Handbook and for identifying some of the issues that are missing. The fact that most of the lines in the different panels are rising to the right is the main reason why inequality is on the agenda. At the same time, the long run of historical data on inequality shows us that there have been periods in the past when inequality fell and when poverty was reduced. Indeed the past 100 years has seen a broadly U-shaped pattern. The series shown in Figure 1 also allow us to underline at the outset the crucial, and often overlooked, point that observed differences in income are not necessarily an indicator of the existence of inequality. Earnings at the top decile (shown in Panel C), for example, may have risen on account of increased costs of acquiring educational qualifications, and not represent any rise in inequality of lifetime incomes. For this reason, it is important to ask what we mean by "inequality."

[1] An alternative colored graph can be found here: http://www.chartbookofeconomicinequality.com/wp-content/uploads/StaticGraphs/USA_staticgraph_coloured.pdf.

[2] We are following here the advice of Schwabish (2014) to avoid "spaghetti" charts.

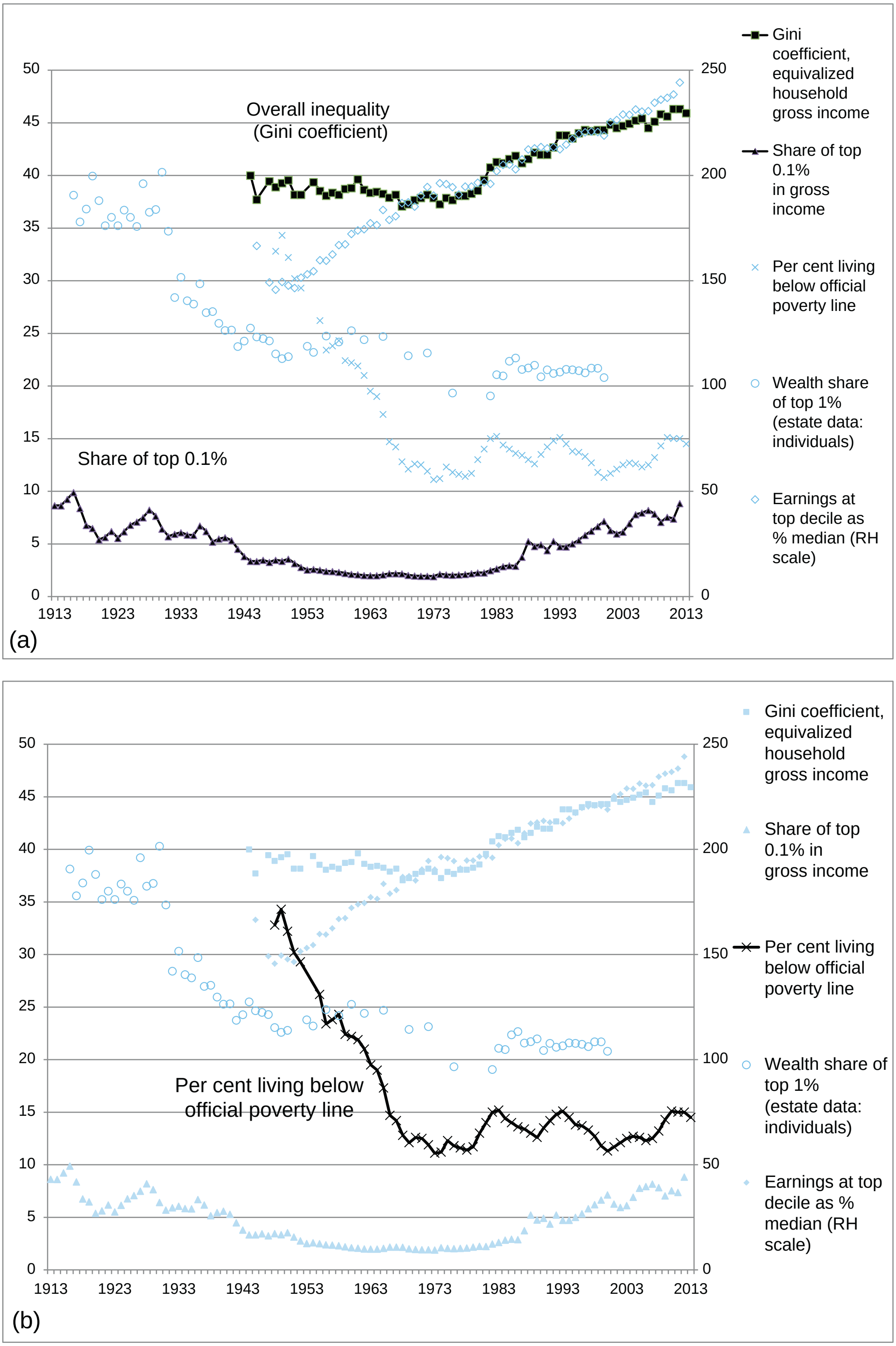
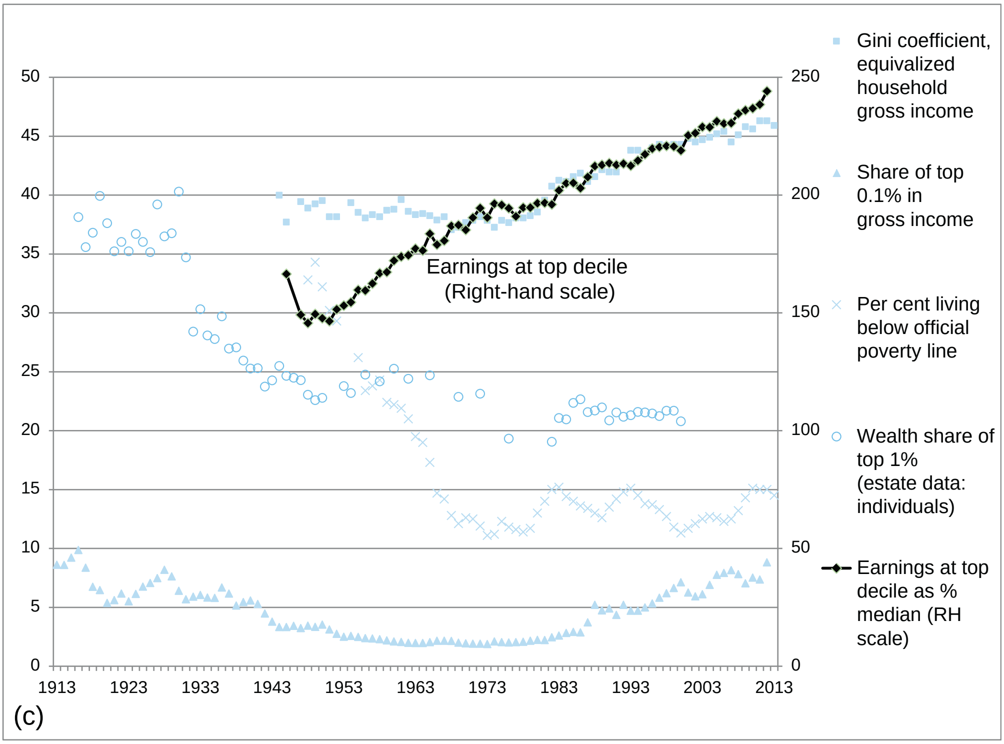


Figure 1 (See legend on next page.)



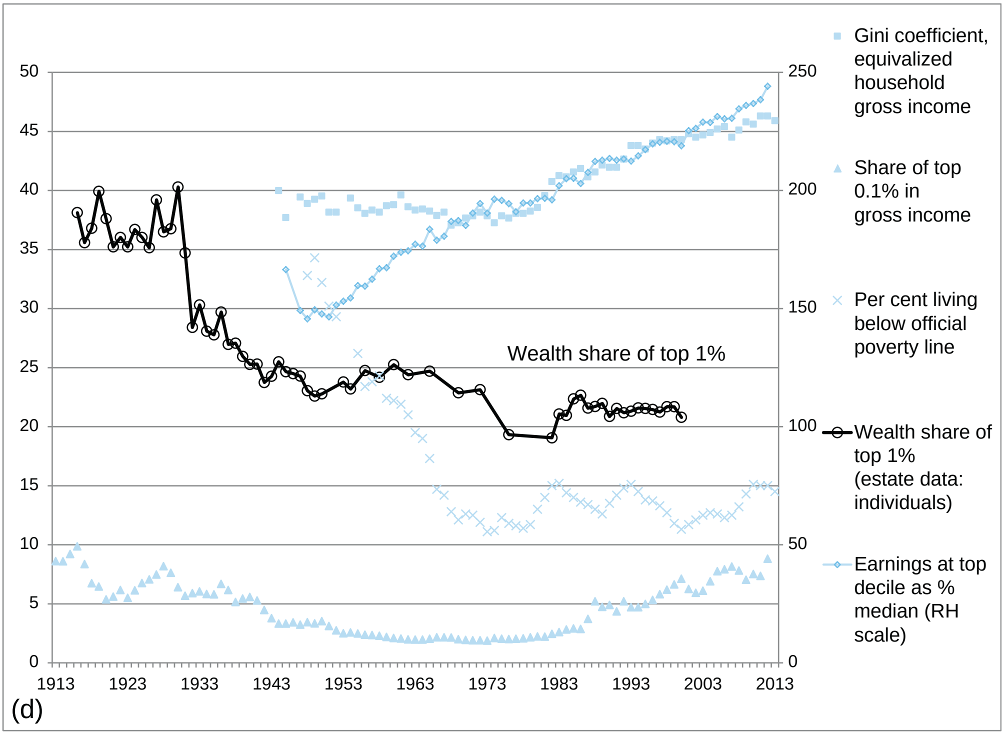


Figure 1 Inequality in the United States from 1913. (a) Overall inequality and top income shares. (b) Official poverty. (c) Earnings dispersion. (d) Top wealth shares. *Note: The right-hand scale relates to earnings at the top decile, and all other series are measured on the left-hand scale.* *Atkinson and Morelli (2014)*.

Box Sources of Data for Figure 1

Overall inequality: Gini coefficient of gross equalized household income from the U.S. Bureau of the Census, Income, Poverty, and Health Insurance Coverage in the United States: 2013 (table A-3, Selected measures of equivalence-adjusted income dispersion), where it has been assumed that half of the recorded change between 1992 and 1993 was due to the change in methods (and therefore 1.15 percentage points have been added to the values from 1992 back to 1967), this series is linked backward at 1967 to the series from 1944 given by Budd (1970, table 6).

Top income shares: The share in total gross income (excluding capital gains) of the top 0.1% is based on the work of Piketty and Saez (2003); updated figures are taken from the Web site of Emmanuel Saez:

http://eml.berkeley.edu/~saez/.

Poverty: The proportion of the population below the official poverty line before 1959 from Fisher (1986) and from 1959 from the U.S. Bureau of the Census Web site, Historical Poverty Tables, Table 2 and Table B1 from the U.S. Bureau of the Census, Income, Poverty, and Health Insurance Coverage in the United States: 2013.

Individual earnings: The series for the top decile of earnings, expressed as a percentage of the median, is based on the Current Population Survey (CPS) from the OECD iLibrary, linked at 1973 to the estimates of Karoly (1992, table 2B.2), and at 1963 to the estimates in Atkinson (2008, table T.10) from the CPS tabulations.

Wealth: The share in total personal wealth of the top 1% of adult individuals is based on estate data from Kopczuk and Saez (2004, table B1).

2. DIFFERENT FACETS OF INEQUALITY

There is much discussion of inequality but there is also much confusion, as the term means different things to different people. Inequality arises in many spheres of human activity. People have unequal political power. People may be unequal before the law. In these two volumes, we are concerned with one particular dimension: economic inequality.

Even limiting attention to economic inequality, there are many interpretations and careful distinctions have to be drawn. It is convenient to first make a distinction between monetary and nonmonetary inequality. The former, refers to standard dollar-valued magnitudes associated with the economic activity of an individual or a household (earnings, income, consumption expenditures, and wealth). Nonmonetary inequality, also referred to here as "beyond income" inequality, addresses broader dimensions of economic life such as well-being or capability.

2.1 Monetary Inequality

Restricting inequality to monetary magnitudes does not prevent confusion. In the media, one often hears statements like "the wealth of the richest *x* billionaires would feed all the poor in a particular country." But this confuses wealth, which is a stock, with income or consumption, which are flows. Flows have to be specified as occurring over a certain period, so that the figures for overall inequality in Figure 1a relate to annual income. Wealth, shown in Figure 1d, is in contrast measured at a point in time. If billionaires gave away all their wealth this year to feed poor families, then they would not appear in the Forbes list next year. If, on the other hand, they gave away the income from their wealth, then the gift would be smaller but they could go on doing it year after year.

There is often confusion between income and earnings. Articles in the academic literature may contain in their titles the words "distribution of income," but they are often actually about the distribution of earnings, and earnings are only part of income. Often too they look only at those in work, and tell nothing about the inequality of income among pensioners or the unemployed. The distinction between earnings and income is made clearly in Chapter 18 by Wiemer Salverda and Daniele Checchi. They observe that there appear to be two largely separate literatures, one concerned with earnings and one with income distribution; their chapter plays an important role in bridging this divide. As they note, it is a question not only of "inequality of what?" but also of "inequality among whom?" Earnings are typically considered on an individual basis. The earnings at the top decile (the person 10% from the top) shown in Figure 1c are those of individual workers, whereas the income measured by overall inequality is the total income of the household.

A person may have zero earnings, and no other income, but live in a household which is comfortably off. Such a situation does of course raise interesting questions—both for the analysis of inequality and in real life. What is the distribution within the household? The curve highlighted in the upper part of Figure 1a refers to the inequality of equivalized household income, which imputes to each member the total income of the household divided by the size of the household corrected by a factor that takes into account economies of scale as well as age-dependent needs. This assumes that all household members enjoy the same well-being. The topic of intrahousehold inequality is addressed in Chapter 16 by Pierre-André Chiappori and Costas Meghir. As they stress, both the level of inequality and the trend over time may be quite different. This issue is particularly relevant to gender inequality, which is the subject of Chapter 12 by Dominique Meurs and Sophie Ponthieux.

Overall inequality is summarized in Figure 1 in terms of the Gini coefficient, and this is the statistic most commonly published by statistical agencies. The typical explanation of the coefficient is geometric: the Gini coefficient is equal to the ratio of the area between the Lorenz curve and the diagonal to the area of the whole triangle under the diagonal. As illustrated in Figure 2, the Lorenz curve shows the proportion of income received by the bottom F percent as a function of F. Where the Lorenz curve is close to the diagonal (curve A in Figure 2), the Gini coefficient is small; where the Lorenz curve hugs the horizontal axis (curve B), the coefficient is closer to 1. If we are comparing two Lorenz

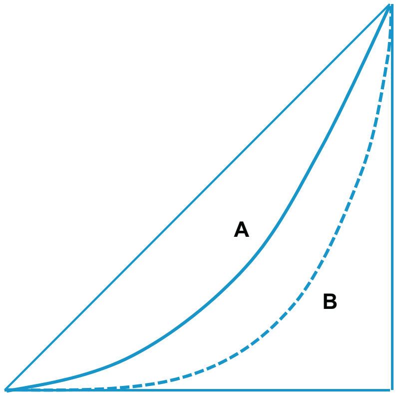


Figure 2 Lorenz curves. *Note: The Gini coefficient of the distribution A is equal to the ratio to the whole triangle of the area between the curve A and the diagonal.*

curves that do not cross, as with A and B, then one (in this case, A) definitely has a lower Gini coefficient. In this case we have Lorenz dominance, and A scores better than B on a wide variety of inequality measures (Atkinson, 1970). The converse is not true. The fact that the Gini coefficient is lower does not imply that the Lorenz curve is everywhere higher: the curves may intersect. The Gini coefficient can also be described in terms of the mean difference. A Gini coefficient of G percent means that, if we take any two households from the population at random, the expected difference is 2G percent times the mean. So that a rise in the Gini coefficient from 30% to 40% implies that the expected difference has gone up from 60% to 80% of the mean. Another useful way of thinking, suggested by Sen (1976), is in terms of "distributionally adjusted" national income, which with the Gini coefficient is (100G) percent of national income. So that a rise in the Gini coefficient from 30% to 40% is equivalent to reducing national income by 14% (i.e., $(100-40)/(100-30)=6/7$ of its previous value).

We may be interested, not just in overall inequality, but also in the top and bottom of the distribution. The top income shares, for which we have the longest run of data in Figure 1, stretching back in the United States to 1913, show the share of total gross income (i.e., before deducting the taxes paid) accruing to the top 0.1%. Figures like these, or the share of the top 1%, have appeared on the placards at demonstrations, such as those of the Occupy Movement. At the bottom of the scale, the poverty figures record the number of people living below the official line, which in the United States dates back to the War on Poverty launched by President Johnson in the 1960s. The evolution of top income shares receives particular attention in Chapter 7 by Jesper Roine and Daniel Waldenström on long-run trends. Chapter 8 by Salvatore Morelli, Timothy Smeeding, and Jeffrey Thompson discusses whether top shares are proxies for overall inequality. This chapter and Chapter 9 by Facundo Alvaredo and Leonardo Gasparini on the post-1970 trends provide evidence about both top and bottom of the scale. In a different sense, the concentration of people at the top and at the bottom of the distribution, or by complement the size of the "middle class," leads to the concept of "polarization," another concept, discussed in Chapter 5 by Jean-Yves Duclos and André-Marie Taptué. There are still many other ways to measure inequality than the Gini or percentage shares, whether it refers to income, earnings, consumption, or wealth. Likewise, there are many ways of expressing social welfare as a combination of some inequality measure and the mean income of a population. These were extensively reviewed in the first volume of this Handbook.

The panels in Figure 1 present therefore quite a rich picture of inequality in the United States over the past 100 years. But there is much that is missing. Figure 1 is a sequence of snapshots, rather than a movie. The top 0.1% in the United States in 1913 were those in the top group in that particular year; some of these people would have fallen out of the group by 1914. The statistics presented in Figure 1 tell us nothing about such mobility, which is the subject of Chapter 10 by Markus Jäntti and Stephen Jenkins. As they explain, there are two aspects. First, there is the subject of how an individual's income changes from one year to the next during their lifetime; on the other hand, there is the subject of income change between generations of parents and children. Such a distinction between intragenerational and intergenerational income mobility reflects the division in the existing literature, but one of the features of their chapter is that it draws out the elements of the measurement of income mobility that are common to both topics. They also raise the measurement issues associated with the inherently multidimensioned information about incomes over time; many of these issues are the counterpart of those that arise when one considers multidimensionality at a point in time (see in the next section).

Figure 1 shows evidence about the distribution of income within one country, but inequality is local and global as well as national. To begin with, money income has different purchasing power depending on local prices, and geographical variation can have a significant impact, as has been shown in the work of Moretti (2013) on the college wage premium in the United States. Spatial inequality is highlighted by Ravi Kanbur in Chapter 20. As he notes, "the spatial dimension of inequality is a key concern in the policy discourse, because it intersects with and interacts with disparities between subnational entities and jurisdictions. These entities sometimes have defined ethnic or linguistic characteristics, and in Federal structures have constitutional identities which naturally lead to a subnational perspective on national inequality." Inequality equally intersects with concerns about globalization, which is the main subject of Chapter 20. The world distribution of income is the subject of Chapter 11 by Sudhir Anand and Paul Segal—see also Bourguignon (2013).

We have referred above to the important topic of gender inequality. This is treated in Chapter 12, where the authors note that much of the literature is concerned with the gender gap in earnings. As they say, this is important, but only part of the story, since there is gender inequality in other forms of income and among those not in paid work, and all these inequalities interact with each other. See also Chapter 20 where Kanbur discusses globalization and gender inequality. Another important missing topic is that of inequality by race and ethnicity. It is a serious omission that we have not included a chapter on discrimination, although we should note that ethnic cleavages are one of the motivations for the theoretical analysis of polarization in Chapter 5.

In introducing considerations such as gender and ethnicity, we are, however, going beyond straight income inequality as pictured in Figure 1. We are abandoning the anonymity that lies behind a Gini coefficient or a top 1% income share. We are asking how unequal income is across various well-identified groups. This is adding a dimension to the inequality concept. Two populations A and B may have the same overall inequality of earnings but the distribution of earnings may be exactly the same for men and women in A, whereas it may be more favorable to men in B. Two countries may have the same share of the top 1%, but in one country they may be all men. World poverty may have fallen, but the number in a particular ethnic group below the poverty line may have risen.

2.2 "Beyond Income" Inequality

Much of the reflection about inequality over the past 15 years or so has been concerned with extending the concept to "beyond income" inequality. As Sen et al. (2010) proposed to incorporate nonmarket dimensions into the measure of social progress, thus going "Beyond GDP," the point here is also to take nonincome dimensions onboard when measuring inequality.

In addition to the quite remarkable perspective it offers on two centuries of economic thought on income distribution and redistribution, the chapter on the history of economic thought (Chapter 1) reminds us that there may be a historical bias in the way economists see and think about income distribution and inequality. This bias is likely to be as present today as it was in the past. Classical economists focused on functional income distribution among land, labor, and capital because they viewed the society they were living in as made up of classes deriving their income from different factors. This view does not fit well our own world, even though factor rewards and the functional income distribution still features today in macroeconomic distribution theories. Another interesting feature of Chapter 1 is the long-lasting reliance in normative economics on utilitarianism. It is only relatively recently, i.e., the 1970s, and very much under the influence of Rawls and Sen, which economists have begun to distance themselves from this approach and to consider alternatives.

It is noteworthy that this line of thought developed very much as a major extension of the previous literature on income inequality measurement, a literature that is comparatively young in a historical perspective. Several features of the income inequality measurement "paradigm" as it developed say between 1920 (Pigou, 1920; Dalton, 1920)[3] and the 1980s are worth stressing as it provides a point of reference for modern researchers who seek to extend this paradigm to a higher level of generality in terms of the definition of economic inequality. A first feature is that most properties of inequality indicators, their interpretation in terms of social welfare, and their analogy with risk analysis are now well understood. To be sure, work remains to be done on the income inequality measurement agenda. This volume of the Handbook gives two examples of it: polarization (Chapter 5) and the introduction of statistical methods in the measurement of income inequality (Chapter 8 by Frank Cowell and Emmanuel Flachaire). But considerable progress has been made. A second feature is that this was done rather quickly after the contributions by Kolm (1966, 1971), Atkinson (1970), and Sen (1973). A third feature is that, despite a possible analogy with utilitarianism, the paradigm was explicitly presented as nonutilitarian, even in the generalized sense of a social welfare function, thus breaking with an important and powerful school of economic thought on social issues. Finally, as the paradigm started to develop, its very relevance was questioned. "Equality of what?" asked Sen (1980), suggesting that the income focus was severely restrictive and even questioning the welfare basis for inequality measurement.

Four chapters in this volume of the Handbook deal in different ways with this extension of the income inequality measurement paradigm. The following simple framework is intended to help identify their main contribution, as well as the general issues in moving from income to other, more general, definitions of economic inequality. We first introduce the concept of "functionings," in the sense of Sen, of individual i, denoted by the vector a_i. Functionings include various aspects of life enjoyment: material consumption, health, job market status, housing and environment quality, etc. Among them, let us single out material consumption, as measured by money income, y_i, so that $a_i = (y_i, x_i)$ where x_i stands for nonmaterial consumption functionings. Let the preferences of individual i among those functionings be described by an ordinal utility $u_i(y_i, x_i)$ function, normalized to unity for some reference bundle $a°$. Let the "satisfaction" of individual i be an increasing function $S[u_i(y_i, x_i), b_i]$ where b_i stands for a set of individual characteristics that can influence the satisfaction that individual i derives from (y_i, x_i). Finally, assume that the economic environment, the technology and social habits are such that the vector of functionings must belong to an individual-specific set Q_i defined by y_i, x_i, and z_i, where z_i is a vector of attributes of individual i, which may differ from b_i. A person may, for example, be able to function at home but not in a formal labor market setting or may come from a social background that broadens her economic opportunities. The set Q_i thus describes the set of functionings individual i can reach, given her characteristics. This may include the standard budget constraint as well as the production function that permits the

[3] They gave their name to the Pigou–Dalton transfer principle, according to which a distribution is less unequal than another if it can be reached by a sequence of mean-preserving equalizing transfers of income.

transformation into functionings of the goods and services bought on the market. Let (y_i^*, x_i^*) be the bundle of functionings preferred by individual i in the set of possibilities, Q_i. Of course both y_i^* and x_i^* are individual-specific functions of z_i. It is also natural to assume that this preferred vector is also the observed vector of functionings. The corresponding satisfaction of individual i is $V_i = S[u_i(y_i^*, x_i^*), b_i]$, where the second argument allows for the fact that two individuals may get different levels of satisfaction from the same functioning bundles.

With the aid of this notation, it is then possible to describe the various approaches in the recent literature to extend the measurement of inequality beyond money income and to see their advantages and limitations.[4] They are discussed in Chapter 2 by Koen Decancq, Marc Fleurbaey, and Eric Schokkaert on inequality and well-being, Chapter 3 by Rolf Aaberge and Andrea Brandolini on multidimensional inequality, and Chapter 4 by John Roemer and Alain Trannoy on inequality of opportunity.

2.2.1 Defining Inequality on Functionings: Multidimensional Inequality Measurement

Inequality is to be defined on the collection of vectors (y_i^*, x_i^*) in the population. Various approaches can then be used with varying degrees of generality. The most obvious way to proceed is to aggregate the various dimensions into a single scalar and apply standard unidimensional inequality measurement to this scalar. The aggregator function, defined as $A(x,y)$, may be arbitrary, satisfying some basic properties, or it may be related to the framework set out in Section 2.2, in which case it is equivalent to assuming that all individuals have the same preferences, u, and the same characteristics b, so that they have the same satisfaction function, $S(u(x_i,y_i), b)$, of a functioning bundle (x_i,y_i). Alternatively, this function of (x_i,y_i) may be the preferences of the social evaluator. In any case, all individuals are assumed to apply the same trade-offs among the various functionings. Such a normative aggregation approach lies behind several specific multidimensional inequality measures based on some functional form for the aggregator function, as in Maasoumi (1986). The new inequality-extended Human Development Index used in the Development Program of the United Nations (UNDP) (Foster et al., 2005; Alkire and Foster, 2010), or the recent efforts by OECD to measure multidimensional living standards taking account of inequality, unemployment, and health (OECD, 2014) are examples of this approach. The recent poverty measurement literature based on the counting of deprivations follows the same logic. Deprivations are grouped by functioning, then the number of deprivations and the extent of deprivation in the various functionings are aggregated to

[4] It also allows us to distinguish the approach adopted in the inequality measurement literature of defining social welfare in terms of incomes y_i from an utilitarian approach where social welfare is defined over u_i (or $S[u_i,b_i]$).

generate an overall poverty indicator for individual i—see Alkire and Foster (2011) and Chapter 3.

Instead of using specific measures and specific aggregator functions, $A(x,y)$ or $S(u(x,y),b)$, it is possible to generalize social welfare dominance analysis in income inequality measurement—the counterpart to Lorenz dominance of income distributions (Kolm, 1977). Following Atkinson and Bourguignon (1982), a partial ordering of distributions of the bundles (y_i^*, x_i^*) in a population may be obtained by considering all the aggregator functions within some class of functions (see, for example, Duclos et al., 2011). As, however, is brought out in Chapter 3 by Aaberge and Brandolini, it is not straightforward to generalize the Pigou–Dalton principle of transfers to the multidimensional case. Moreover, we should not lose sight of the fact that nonmaterial functionings may require different treatment. The choice of aggregator function has to reflect the specific characteristics of, say, health as a variable. For example, health may be measured in ordinal rather than cardinal form, as is discussed by Allison and Foster (2004) and Duclos and Echevin (2011).

2.2.2 Individual Preferences and the Income Equivalent Approach

Interpreted in terms of individual preferences, the preceding approach imposes identical preferences on all individuals. If ordinal preferences are observed, then a particular individual-specific aggregator function can be used which has the dimension of income—see chapter 4 of Fleurbaey and Blanchet (2013). With a reference vector x° of nonincome functionings, the equivalent income corresponding to an observed bundle (y_i,x_i) is given by the solution $\mathring{y}_i$ of the equation: $u_i(\mathring{y}_i, x^\circ) = u_i(y_i, x_i)$. Such a solution does exist if $u_i(\ )$ is reasonably assumed to be continuous and monotonically increasing. The equivalent income is thus a function of the bundle (y_i,x_i) conditional on x°. It can be handled exactly in the same way as income is dealt with in income inequality measures. Of course, the issue is whether it is possible to estimate individual preferences $u_i(y,x)$. In Decancq et al. (2014), this is done using subjective satisfaction data and relating these data to income and observed functionings for individuals with common characteristics in some subset of (b_i). Another issue is the choice of the reference nonincome bundle, x°, since inequality measures will be conditional on that bundle—see Chapter 2.

2.2.3 Defining Inequality Using Subjective Satisfaction

Going further along the chain, another approach at measuring inequality is to focus directly on satisfaction levels, $V_i = S[u_i(y_i^*, x_i^*), b_i]$, as directly observed in satisfaction surveys. This is equivalent to using the individual-specific aggregator function, $u_i(y_i^*, x_i^*)$, as well as the satisfaction function $S(u,b_i)$. Several authors have followed this route, showing for instance that, unlike for observed income, inequality of "happiness" tended to go down over time in most developed countries and, as for income, to be smaller in richer than in poorer countries—see for instance Veenhoven

(2005), Stevenson and Wolfers (2008) on the United States, Ovaska and Takashima (2010), Becchetti et al. (2011), and Clark et al. (2012).

Since life satisfaction is generally recorded on an ordinal scale, there are evident problems in converting such measures to a cardinal scale and employing standard inequality indicators. For this reason, an alternative approach based on dominance criteria has been proposed by Dutta and Foster (2013). On the conceptual side, the issue arises as to the interpretation to be given to life satisfaction data. In particular, they may actually reveal the satisfaction of an individual relatively to his or her past experience and future expectations, and also possibly in relation to other individuals. If so they are improper for the measurement of "economic inequality." In terms of the notation introduced above, it is doubtful whether the individual characteristics b_i, which determine how ordinal preferences are cardinalized into satisfaction, should be taken into account in measuring economic inequality. The same bundle of functionings enjoyed by two individuals having the same preferences may yield different levels of satisfaction if one of them has a rather positive and the other a negative attitude toward life in general. A more detailed account of the interpretation to be given to subjective satisfaction data is offered in Chapter 13 by Andrew Clark and Conchita d'Ambrosio.

2.2.4 Inequality of Capabilities

Instead of defining inequality on the observed bundle of functionings (y_i^*, x_i^*), resulting from the choice of individual i in his/her choice set, Q_i, one could move upstream and consider the inequality of these choice sets, as in the capability approach. The first step is to identify these sets, as different from the particular point (y_i^*, x_i^*) actually achieved in the set. Presumably, this could be done by considering the set generated by all the functioning bundles observed for persons with the same vector as attributes, z. Assuming that such identification has been made, the second step would consist of defining an inequality measure on these sets rather than scalars, as with income, or vectors as with multidimensional inequality. In practice, the measurement of inequality based on capabilities has been reduced to emphasizing a few components of the vector of individual characteristics, z_i, that indirectly define the size of the set Q_i. Typically, these are variables like education, health, or the availability of material resources. These three variables are combined linearly at the aggregate national level to define the familiar Human Development Index used by the UNDP to measure the inequality of functionings between any pair of countries. Several attempts have been made to incorporate a larger set of variables at the individual level in specific countries (see, for instance, Anand et al. (2007).

2.2.5 Inequality of Opportunities

Inequality of opportunity is defined as that part of the inequality in optimal functionings, (y_i^*, x_i^*), that is, due to differences in individual characteristics, z_i, or possibly in a subset of these variables. Among these characteristics, a distinction is made between variables, on the one hand, that are outside the control of individuals, which are called the "circumstances" faced by the individual, and variables, on the other hand, which may be assumed under his/her control. Then the inequality of opportunity can be measured by the inequality that exists in the (y,x) space across groups of people facing the same circumstances. Those groups are called "types" by Roemer (1998). The simplest measure is obtained by considering the inequality of the mean vector (y, x) among the various types. More elaborate measures would compare the distributions (y,x), rather than means, across types. In Chapter 4, Roemer and Trannoy discuss methods for doing so when focusing on income distribution only. The implementation of this approach requires assumptions to be made about what may be considered as "circumstances" (gender, ethnicity, family background, . . .) and what may be assumed under the control of an individual (school achievement, for instance). Moreover, there are necessarily unobserved circumstances, so that one can at best measure the income inequality due to observed circumstances. A particular case frequently used is when one considers only one component of z_i, like ethnicity. In the space of earnings, the inequality of opportunity then corresponds to the familiar concept of wage discrimination, as is discussed for gender in Chapter 12.

Inequality in terms of opportunity or capability is both defined on an "*ex ante* basis," that is among groups of people with some common characteristics, and irrespective of differences in individual achievements in the space of functionings. (In contrast, inequality of outcomes, be they income, earnings, consumption spending or even "satisfaction," is an *ex post* inequality concept.) Practically, the main difference between the measurement of the inequality of capability and opportunity is that the former focuses on the inequality in the vectors of attributes, z_i, whereas the simplest approach to the latter considers the inequality between mean outcomes (y,x) across various "types" defined by identical "circumstances," z. A second difference may lie in the individual characteristics selected to differentiate people, determinants of the set of possible functionings on the side of capabilities, and exclusively circumstances outside the control of individuals on the side of opportunity. Recently, there have been a number of attempts at measuring the inequality of opportunity in several countries—for instance, Brunori et al. (2013) have put together estimates obtained for a set of 40 countries using earnings as outcome and family background as circumstances. A related literature, although with more antecedents, is concerned with intergenerational income and more generally social mobility—see Chapter 10. Here too, the issue is to measure the contribution of an observed circumstance, the parents' social status, the inequality of a particular functioning, and the children's social status.

Overall, "beyond income" inequality measurement touches upon the very fundamentals of economic inequality and a number of important conceptual advances have been made since the pathbreaking work by Rawls and Sen. Data limitations, the frequent difficulty of translating conceptual parameters into actual figures, and quite often the complexity of the analytical instruments being proposed have limited the empirical applications, but the area is promising for future research. A priority should be the search for simpler ways of making use of the conceptual advances in describing the multiple dimensions of economic inequality without reducing them to a single number.

3. DATA ON INEQUALITY

We began this chapter with the concrete example of Figure 1 in order to highlight the centrality of data. In all fields of economics, data play a key role, but in a field as politically charged as inequality, it is especially important to be careful with the quality of data. When faced with charts like those in Figure 1, showing the evolution of inequality, one should not take them simply on trust. One should ask: what data are there? Where do they come from? Are they fit for purpose? In what follows, we concentrate on data on monetary inequality, particularly income, but similar questions arise with data on non-monetary variables such as material deprivation or happiness.

3.1 Care with Data

There are two dangers. The first is the inappropriate use of data. All too often, people make claims that inequality is increasing, or decreasing, on the basis of comparing data at two different dates that are not comparable. Or country A is claimed to be performing better than country B on the basis of statistics derived from sources that cannot be compared. The share of the top 10% in total wealth may be obtained in one country from a household survey and in the other from the records of administering a wealth tax. The other danger is that of going to the other extreme and rejecting all evidence about inequality on the grounds and that it can only be measured imperfectly. That is a counsel of despair.

In our view, all forms of possible evidence should be brought to bear, but we need to take full account of their strengths and of their weaknesses. Here there have been remarkable advances. When one of us (ABA) started research on poverty in Britain in the late 1960s, the British government had decided not to allow access to the household records. The only materials were published tabulations. This changed in the 1970s. The first volume of the Handbook, published in 2000, could draw on the household survey data that had become much more widely available. Not only were there many more surveys being conducted, typically by statistical agencies, but also researchers more commonly had access to the microdata. Although this change is far from universal, it has allowed scholars to assemble internationally comparable datasets, notably the Luxembourg Income Study (LIS) founded by Tim Smeeding, Lee Rainwater, and Gaston Schaber in 1983, and the World Bank's PovcalNet covering some 850 household surveys from 127 countries.

In the decade and a half since the first volume, there have been at least four major departures. The first is the rapid growth of experimental research in economics, represented here by Chapter 13 by Andrew Clark and Conchita d'Ambrosio, where the authors show how data generated in experiments, together with survey evidence, can throw light on the subtleties of attitudes to inequality. The second is the much greater access to distributional data from administrative records. When the European Union Statistics on Income and Living Conditions (EU-SILC) replaced the European Community Household Panel from 2003, the regulations allowed flexibility as to the source of data and increasingly Member States have drawn information from administrative sources. The third is the renewed interest in historical data. Inspired by Piketty (2001), there has been a considerable investment in the construction of long-run time series, notably covering top income shares, as made available in the World Top Incomes Database administered by Facundo Alvaredo. Finally, the fourth improvement is the increasing standardization of the collection of data across countries, which permits more rigorous comparative work. EU-SILC is an example in Europe; the Program for the Improvement of Surveys and the Measurement of Living Conditions in Latin America and the Caribbean is another example. Even though more efforts are needed, cross-country comparisons today definitely make more sense than was the case one or two decades ago.

These developments mean that we now are much better informed about the extent of economic inequality and the trends over time, as is clear from reading the surveys of the evidence in Part II. The historical research is examined in Chapter 7 on long-run trends in the distributions of income and wealth, covering more than 25 countries and going back in some cases to the eighteenth century. The post-1970 evolution is the subject of Chapters 8 and 9. The former covers inequality and poverty in OECD and Middle Income countries, demonstrating that "data have come a long way" since the chapter by Gottschalk and Smeeding (2000) in volume I. The latter covers developing countries, where again there has been great progress in the measurement of inequality and poverty. Chapter 11 investigates the world distribution of income and global poverty.

The availability of data is, moreover, one key ingredient in the study of the causes of economic inequality, which is the focus of Part III. In many cases, these investigations are based on statistical analysis of country panel datasets on inequality (derived by pooling time series of observations for each of a number of countries). Differences over time and differences across countries are used to explore the multiple causes of inequality. This is the explicit concern of Chapter 19 by Michael Förster and István Tóth, and it underlies much of the analysis of Chapter 18 on the distribution of earnings by Wiemer Salverda and Daniele Checchi. In Chapter 21, the econometric analysis of the relationship among democracy, redistribution, and inequality by Daron Acemoglu, Suresh Naidu, Pascual Restrepo, and James Robinson uses international databases on inequality. It is on such international databases, now widely used, that we concentrate here, since they illustrate many of the issues.

3.2 International Databases on Income Inequality

In considering international income distribution databases, as listed and discussed in Chapter 19, a key distinction is between primary databases that rely directly on micro-data, standardized as much as possible to ensure comparability across countries and time periods, and secondary databases that compile estimates of income distribution indicators from available published sources. Examples of the former include LIS, the EU-SILC—which also coordinates the data collection in the countries covered—the OECD income distribution database, SEDLAC covering Latin America and the Caribbean and the World Bank's POVCAL/WYD. Secondary databases include the World Income Inequality Database (WIID) assembled by UNU-WIDER (an updated version of the dataset originally constructed by Deininger and Squire (1996) at the World Bank), and the "All the Ginis you ever wanted" database put together by Branko Milanovic (2013), also at the World Bank. The latter states clearly that "this dataset consists only of the Gini coefficients that have been calculated from actual household surveys," and a second important distinction is between databases, like "All the Ginis," that are restricted to actual observations and those databases that impute missing values of specific indicators for some countries and for some time periods. Aiming at "the widest possible coverage across countries and over time," the Standardized World Income Inequality Database (SWIID) "uses a custom missing-data algorithm to standardize the United Nations University's WIID and many, many additional observations" (SWIID Web site and documentation). The University of Texas Inequality Project (UTIP) Estimated Household Income Inequality Data Set is derived from the econometric relationship between the UTIP-UNIDO dataset on industrial pay, other conditioning variables, and the World Bank's Deininger-Squire dataset on income inequality.

The original World Bank Deininger and Squire (1996) database was scrutinized by Atkinson and Brandolini (2001), who showed the risks of reaching inconsistent conclusions using secondary databases, or, more generally, comparing income distribution indicators across countries or time periods that relied on different definitions of income or the statistical unit. While recognizing the value of such databases, they cautioned against their uncritical use and set out a number of principles which should guide the construction of secondary databases. Progress has been made since then. In a paper reviewing the WIID database for a special issue of the *Journal of Economic Inequality*, edited by Ferreira and Lustig (2015), on income inequality databases, Jenkins (2014) repeated the comparison made by Atkinson and Brandolini between the database and consistent estimates obtained from LIS for a sample of rich countries in the early 1990s, using WIID version 2c (2008) and found that differences had been reduced (and a new WIID version 3.0B was subsequently released in 2014). Yet, he reiterated that "one cannot simply use the WIID data 'as is'" (Jenkins, 2014, p. 15). As this kind of benchmarking has not been made for developing countries, it is not unlikely that inconsistencies are more frequent there. One should not use data from secondary databases without first making a careful inspection.

3.3 A Checklist of Questions

When using data on income inequality, what questions should one ask? Here, we give a checklist covering some of the most important. These issues, and many others, are discussed in the Canberra Group Handbook on Household Income Statistics (United Nations Economic Commission for Europe, 2011), which is the second edition of a handbook produced in 2001 by an International Expert Group on Household Income Statistics established in 1996 at the initiative of the Australian Bureau of Statistics. In describing it as a "checklist," we are not suggesting that there is a single right answer. The appropriate choice depends on the context and may differ between countries at different stages of development. The choice depends on the purpose of the analysis. But it is essential that the user be aware of what data they are employing.

3.3.1 Inequality of What?

In some countries statistical offices collect data on household income, whereas in others consumption expenditure data are collected. The Povcal database comprises both countries that report income inequality and others that report inequality in consumption expenditure. LIS avoids this heterogeneity by using income surveys for all countries. Income may be defined in a variety of ways: posttax (or disposable) income, pretax income allowing for deductions, such as interest paid (confusingly, this is often called "net income" in official statistics), or pretax income before deductions. As implemented, the income concept may follow more or less closely the definition adopted by the International Conference of Labour Statisticians (and the second edition of the Canberra Handbook), which covers all receipts whether in monetary form or in kind, apart from irregular or windfall receipts. Important issues here (as for the definition of consumption) are the inclusion or exclusion of imputed rent for owner-occupied housing, of home production, and of in-kind benefits.

Income and expenditure relate typically to a year, but may be measured over different time periods. This is particularly important in the case of earnings, as discussed in Chapter 18. The reference period may be the latest pay period or earnings that may relate to normal monthly earnings, excluding irregular bonuses, or they may be total annual earnings. They may be expressed per hour, and this may allow a decomposition into wage and leisure inequalities. The issue of timing also affects the population covered. People may be present for part of the year, and the inclusion or exclusion of such part-year incomes, or earnings, affects the measured degree of inequality. Another issue related to the comparability of earnings data across countries is the status of payroll charges and social security contributions. Earnings are net of all these charges in some cases and gross of contributions paid by employees in other cases, whereas payroll charges paid by employers are rarely recorded. From that point of view, progress in constructing international databases on income distribution has not been paralleled by the same effort for individual earnings.

An additional issue of importance with existing datasets on income inequality is the difference in the cost of living across geographical areas in the same country. Such data do not exist in all countries. Yet, these differences may be sizable and may have a major impact on the estimation of inequality. Uncertainty about the rural–urban cost of living differential has led the managers of the Povcal database in the World Bank to report separately on the distribution of income in rural and urban areas in both China and India and differences in reported estimates of income inequality in China are often due to different assumptions about the rural/urban cost of living ratio. Differences in the cost of living across cities—if only with respect to housing rents—in developed countries generate the same kind of imprecision (we have already cited the work of Moretti, 2013).

3.3.2 Among Whom?

Data on inequality may refer to differences between households, between inner families, between tax units, or between individuals. Much empirical evidence relates to households, and surveys are typically conducted on this basis. Such a measure however tells us nothing about the distribution within the household—the subject of Chapter 16. Where there are several generations of adults within the household, inequality may be concealed. The same applies to the inner family, in that the aggregation of the income (or consumption) of a couple conceals gender inequality—the subject of Chapter 12. In this context, it is interesting that a number of countries have moved to individual taxation under the personal income tax. From such administrative data, we can learn, for example, that women were seriously underrepresented among the top 1%. In Canada, in 2010, women accounted for only 21% of those with gross incomes in the top 1% (Statistics Canada, 2013); in the United Kingdom, in 2011, the corresponding figure was 17% (Atkinson et al., forthcoming).

Households and other units have differing size and composition, and adjustments have to be made using equivalence scales. Here, there is a variety of practice and some harmonization across primary databases would be welcome. For instance, the "equivalization" procedure differs between LIS ("Key Figures" webpage) and the OECD income distribution database, which use the square root of the total family size as an equivalence scale, and Povcal which uses the total family size, and imputes the total household per capita to each household member. In other words, the equivalence elasticity is set to 0.5 in the first case and to 1 in the second. These choices can make developing and emerging countries appear more unequal in comparison with developed countries, were the definition of income the same in both groups. To reestablish comparability, it would not be difficult for all the databases to provide estimates of income distribution indicators with both equivalence elasticities—something that is done in the SEDLAC database for Latin America. At the same time, it is not clear that economies of scale in consumption and therefore the equivalence scale should be the same across countries at different development levels.

Except where the reference unit is the individual, there is the further question of the weighting of observations—an issue that is often neglected, and not always documented. If we observe income at the household level, it does not follow that each household should be regarded as one unit regardless of size. Weighting is a separate issue from the choice of equivalence scale. The income of a household may be corrected by an equivalence scale that allows for economies of scale within the household but that does not mean that it should be weighted by the number of equivalent units. Weighting by the number of individuals may be judged more appropriate. This has, of course, the consequence that total income attributed to the multiperson households is greater.

3.3.3 Data Sources

Each source of data has strengths and weaknesses. Historically, evidence on income inequality, such as that used by Kuznets (1955), came from administrative records, of which the most important were the statistics derived from personal income taxation. The income tax data have serious limitations—the incomplete coverage of those below the tax threshold, the underreporting of income, and the impact of lawful tax avoidance and income shifting—which are discussed extensively in Chapters 7 to 9. They must therefore be used with caution. The same applies to the source that is now more widely used: household surveys.

In the case of household surveys, differences in survey questionnaires and in the methodology of correcting for nonresponse or missing observation reduce the comparability of inequality indicators. In his review of LIS, Ravallion (2014) emphasizes the issue of nonresponse and missing income data. Nonresponse by sampled households is in some cases handled by redrawing a comparable household in the same stratum, and in other cases by simply reweighting responding households. But there is a risk of a bias if nonresponse is relatively frequent and not random with regards to income. This bias is likely to be substantial if the very top of the distribution is simply not sampled, as it is often the case in developing countries—as shown by Korinek et al. (2006). The frequency of nonresponse might usefully be reported by statistical offices. The same applies to missing income values for responding households. In some cases, a value for total income or for an income component is imputed based on observed characteristics of households and household members. In others, no correction is made with the effect of the corresponding observation being taken out of the sample on which the income distribution is estimated. Again, in both cases, there is a clear problem if missing values are income-dependent.

3.3.4 Relation with National Accounts

Data from administrative records or from household surveys have to be viewed in relation to the national income accounts, which provide an important point of reference. Indeed, in the case of income tax data covering only part of the population, the national accounts are the standard source of the independent income control totals. In the case of household surveys, the issue may arise at the level of total household income, as has been extensively discussed in the literature on world income inequality—see Chapter 11. It may arise when some average income component in a survey appears to be relatively more underestimated in comparison with National Accounts than other components, provided that a full-fledged household account is available. In some cases, the statistical office scales up that income component so as to establish consistency with the National Accounts total. But where this is not distributed as total income, or where the discrepancy is due to underreporting as well as nonreporting, this may drastically modify most income distribution indicators. This kind of correction is now rarely done in advanced countries, but is still applied in some emerging countries, especially for property income, most often grossly underreported in household surveys. The database managed by the Economic Commission for Latin America and the Caribbean includes such an adjustment. In Chile, for instance, all income components in the CASEN survey (salaries, self-employment income, property income, transfers, and imputed rents) are scaled up (or down in the case of imputed rent) so as to match National Accounts. The only exception is for property income, for which the gap is imputed entirely to the top quintile of the distribution. As there are differences in the definition of income in surveys and in National Accounts, such a correction introduces additional noise in the distribution data that appear later in cross-country databases. Bourguignon (2014) indeed shows that the size of the adjustment may be substantial.

3.4 Implications of Data Heterogeneity

The consequences of the heterogeneity of income distributions indicators in cross-country databases for economic analysis and policy are important. In the first place, they make benchmarking across countries or time periods a fuzzy exercise. Not being able to check unambiguously that inequality has increased or decreased in a given country or to compare such an evolution to what has occurred in neighboring countries is a serious handicap for policy making and for the democratic debate in general. Relying on the most transparent and comparable measurement apparatus of income inequality is absolutely essential.

A second consequence of the imprecision and the lack of comparability of income distribution indicators is the weakening of standard econometric analyses of the consequences of income inequality. A noisy regressor introduces a bias in any regression. At the limit, if the noise is too big the estimated coefficient of that regressor goes to zero and income distribution is deemed unimportant in explaining, for instance, the pace of economic growth, political instability, or crime. Consider, for example, the widely cited study of Ostry et al. (2014), who test the influence of inequality and the extent of redistribution, as measured by the difference between the inequality of gross and net incomes, on growth. They find that “lower net inequality is robustly correlated with faster and more durable growth, for a given level of redistribution [and that] redistribution appears generally benign in terms of its impact on growth; only in extreme cases is there some evidence that it may have direct negative effects on growth. Thus, the combined direct and indirect effects of redistribution—including the growth effects of the resulting lower inequality—are on average pro-growth" (2014, p. 4). As the authors clearly recognize, these conclusions must be taken with care if one expects substantial measurement error in the difference between net and gross income inequality, and one should therefore look at the underlying source.[5] The study makes use of version 3.1 of the SWIID database created by Solt (2009). The SWIID data are indeed extensive, providing Gini coefficients for gross and net income for some 175 countries for years in the period 1960–2010. There are more than 4500 observations in the version 3.1 of the database. At the same time, many of these observations are imputed: many countries in the sample lack regular observation of both gross and net income inequality.[6] The author of the SWID database is to be commended for providing the standard deviations of imputed values, but a two standard deviation range for the Gini coefficient in Bhutan, for example, in 2012 of 24–45% (from SWIID version 4.0) means that there is limited information content, as does the range for Malaysia in 2012 of 32–61%. This means in turn that users need to take account of the underlying data quality and that studies that fail to do so are open to question, as is emphasized by Solt (2009). (Version 4.0 of the SWIID is set up to facilitate the application of the multiple imputation approach to parameter estimation, and we understand that version 5.0 goes further in that direction.)[7]

Another consequence of the inconsistency of the income distribution indicators found in cross-country databases is the likely inaccuracy of global income distribution indicators, which cumulate the measurement errors to be found in national income distribution indicators. Global income inequality estimates are certainly extremely noisy, as suggested by the discussion in Chapter 11, although the imprecision of national income distribution indicators is only one part of the problem. The inequality between countries represents a high share of total global inequality so that another major source of ambiguity lies in the estimates of the mean income of national populations relatively to each other. In both cases, moreover, it is clear that big countries play a major role, whereas the

[5] Chapter 21 by Acemoglu et al. in this volume uses the same database on income inequality with two important differences. On the one hand, they use only net income inequality rather than the difference between net and gross, necessarily more imprecise. On the other hand, the inequality appears on the left-hand side of their regression equation so that measurement errors do not create estimation biases.

[6] SWIID provides estimates of redistribution in countries where the source data include at least three observations of net inequality and at least three observations of market inequality (and excludes some countries that meet this criterion but the two are not contemporaneous). Ostry et al. (2014) also provide estimates of the relationship among growth, inequality, and redistribution restricted to that sample.

[7] Such improvements in datasets are much to be welcomed. At the same time, they mean that we should revisit the conclusions drawn in studies based on earlier versions. It also underlines the importance of recording the version of the dataset used and of maintaining archives.

imprecision on the degree of inequality or the mean income within small countries has very little impact on global inequality. This is well illustrated in Chapter 11 where the difference in the rate of growth of the mean individual standard of living in India as reported in the household surveys and in National Accounts is shown to have a significant impact on trends in global inequality. More work is needed to evaluate the degree of imprecision of global inequality estimates due to these different causes—imprecision of national income distribution data, of national means, and of course purchasing power parity estimates—so as to estimate confidence intervals and being able to check whether or not estimated changes are significant. The same applies to global poverty measurement, in particular when it is defined on the basis of a common absolute poverty line.

3.5 The Way Forward

What is the way forward? How can we improve our ability to make international comparisons of distribution and distribution trends, whether for benchmarking, econometric analysis, or global distribution estimates? In comparing income distribution across countries and over time, one would ideally like to access microdata and compute the appropriate summary measures controlling for the definition of the income unit (household or individual), income (gross, net, consumption expenditures, including or not in kind transfers, imputed rents, . . .). But, of course, this would be a herculean task. Hence, there is an obvious need for a first treatment of the data done once for all by the database managers rather than by every user of the database. This requires that data be standardized, as much as possible, in agreement with some consensual definition of income and income units. The "Key Figures" on the LIS Web site obey that logic, while access to the original microdata (and to the STATA or SPSS programs that generate the Key Figures) allows users to depart from this core definition.

Progress in this direction can best be made by following the route of national accounts, with the analogue of the UN System of National Accounts being developed, building on the work of the Canberra Group on Household Income Statistics and on regional initiatives such as the European Union social indicators (Atkinson and Marlier, 2010). Guidelines could then be agreed for the assembly and analysis of distributional data. But in the case of income distribution analysis, a further step is necessary, since a key element is that of access to the microdata. What is required is the possibility for outsiders to access the microdata themselves, under conditions that guarantee confidentiality as in LIS. The same kind of architecture could be developed in other regions or possibly within an international institution like the World Bank. Such guidelines and agreed access to income distribution microdata would not however solve the data problems inherent to the unavoidable incompleteness of the surveys. Moreover, these problems, notably those of securing adequate response rates, may in the future become more severe. From that point of view, complementing standard survey-based analysis with administrative sources has proved to be extremely promising in the recent years.

A number of European countries have moved in this direction, with their EU-SILC data being collected in this manner (although this does raise issues of comparability with the data from countries that rely solely on household surveys). The use being made of the top income database based on tax data in developed countries is a sign of the importance of complementary data—see the discussion in Chapter 8. Combining both sources is not an easy task. The reference unit is not always the same, household in one case, tax unit in the other. Income concepts may differ across the two types of source. Moreover, it is not clear whether top income individuals are absent from household surveys—the nonresponse issue alluded to above—or whether they are present but with underreported income. The correction to be made to inequality indicators is not the same in the two cases—see Alvaredo and Londoño Velez (2013). More generally, the required adjustments may differ across countries.

In line with the discussion in Section 2, one may also wonder whether cross-country inequality databases should not go "beyond income" and incorporate other dimensions relevant to economic inequality. Without getting into the difficulty of measuring the inequality of capabilities or opportunities, some components of a broader definition of inequality can easily be measured, and the dimensions extended. This is the case, in particular, of inequality across gender. Such a database, based on Labour Force Surveys, does exist for OECD countries at the OECD and also for some emerging countries in LIS' "Key Figures." It should not be a major effort for most other primary databases relying on household or labor force surveys covering a larger set of countries to report summary statistics on gender earnings ratios. More generally, primary databases could try to go "beyond income" by reporting summary statistics on the joint distribution of income variables and other individual or household attributes available in standard surveys. Education, gender, and ethnicity are the most obvious examples.

Going beyond the exploitation of standard income focused surveys is problematic because relevant attributes are typically covered in different surveys. For instance, the Demographic and Health Surveys (DHS) in developing countries cover self-reported health status, fertility, infant mortality at the individual or household level. Yet, they do not collect direct information on monetary resources, so that nonincome functionings cannot be considered jointly with income. Matching techniques with household income or consumption surveys could be used to impute an income to households in the DHS but then it is difficult to deal with the inherent imprecision. On the other hand, there are numerous international databases that combine income inequality data with other dimensions of functionings. In the field of health, this was achieved by the Globalization-Health Nexus database put together by Cornia et al. (2008). The problem with these databases, however, is that the nonincome indicators are essentially aggregate so that those databases generally give no information on the inequality of the corresponding nonincome attributes, and, of course, still less on their joint distribution with other attributes, including income or earnings. From that point of view, generalizing and standardizing poverty surveys that include questions on various types of deprivation may be the simplest way of monitoring one aspect of "beyond income" inequality.

4. TAKING ECONOMIC THEORY SERIOUSLY

Data are one important ingredient in studying economic inequality; the second important ingredient is provided by economic theory, which underpins the search for explanations of inequality in Part III. Here again, we make a plea for taking seriously the building blocks that are utilized. One cannot simply take an economic model off the shelf and apply it in an unthinking way to the problem at hand. Likewise, theorists must keep an attentive eye on basic empirical facts and make sure their representation of the way multiple economic mechanisms combine to generate specific properties of the degree of inequality in an economy fits these facts. In what follows, we illustrate and discuss this twofold requirement for identifying the mechanisms that perpetuate or modify inequality by focusing on the role of technical progress, human capital, and wealth accumulation within a largely macroeconomic framework. As can be seen from Chapters 15 and 16, models that incorporate these particular mechanisms are indeed central in the present theoretical reflections on economic inequality, including the accent recently put on the top of the income and wealth distribution. Much of the recent reflection on the possible causes of the observed rise in inequality actually bears upon macroeconomic factors.

In this section, we focus on the determinants of market incomes: wages and capital incomes. These subjects, particularly wages, are covered extensively in other Handbooks, such as the *Handbook of Labour Economics*, and in designing this volume we have sought to avoid overlap. For the same reason, we have devoted more space to this aspect in this Introduction, as a contribution to bridge-building across fields of economics. We should also underline that while wages and the return to capital are important elements in determining the distribution of household income, their impact depends on a variety of social and institutional mechanisms, such as household formation and demographics, and on the redistributive incidence of public policy (see Section 5).

4.1 The Race Between Technology/Globalization and Education

In the Introduction to volume I, we set out the application of supply and demand analysis—perhaps the simplest of economic theories—to the explanation of rising earnings dispersion. Jan Tinbergen (1975) famously described a "race" between increased demand for educated workers and the expansion of the educated population. Where demand—driven by new technology or by globalization—outstrips supply, then the premium for education rises. As typically portrayed, as in Figure 3, the supply and demand equilibrium is shifting up over time. The wage premium for higher-educated workers is rising because technological progress is biased in their favor—the skill-bias technical change (SBTC) hypothesis—or because increased global competition favors more

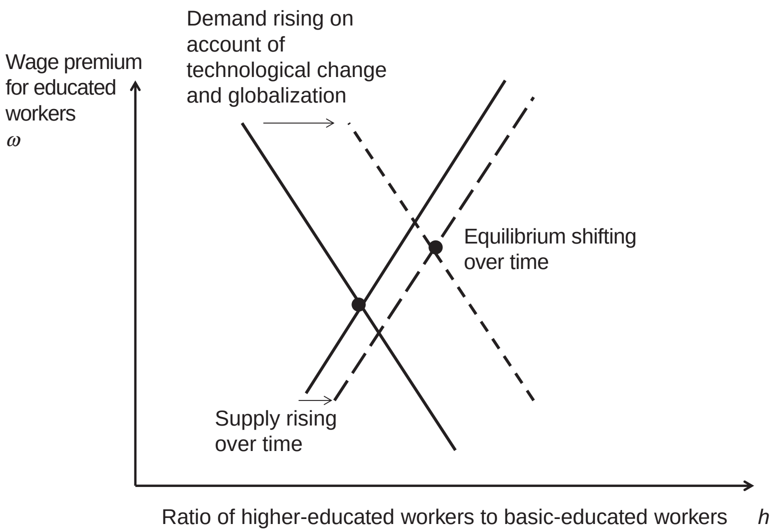


Figure 3 The "race" between technology/globalization and education.

educated workers. In what follows, we refer mainly to the SBTC hypothesis, but this does not mean that we discount the role of international trade.

First year economics appears to explain what is observed in the real world. However, second year economics teaches us that a race is a dynamic process and its outcome depends on how one specifies the underlying adjustments. Suppose, as seems reasonable, that at any moment, t, the ratio of higher-educated to basic-educated workers is fixed at $h(t)$ and that the relative wage, ω, clears the labor market. With aggregate output a function of the two kinds of labor, with a constant elasticity of substitution σ, the wage premium is determined by

$$\omega = A\,[\alpha_h/\alpha_b]^{(1-1/\sigma)}h^{-1/\sigma} \tag{1}$$

where A is a constant and α_i denotes the productivity of workers of type i (h for higher-educated and b for basic-educated). If over time skill-biased technical change (SBTC) raises the square bracket, and if (a condition that is often forgotten) the elasticity σ is greater than 1, then the wage premium rises for any given $h(t)$. In general, however, h will rise in response to the rising premium. If the growth rate of a variable x is denoted by $G(x)$, and the growth rate of the square bracket is a constant, g, then we can write

$$G(\omega) = (1-(1/\sigma))g - (1/\sigma)G(h) \tag{2}$$

Suppose that the growth rate of h responds to the difference between the wage premium and the cost of acquiring education with an elasticity β; moreover, suppose that the cost of education, in terms of the wage of basic-educated workers, is equal to a fee, F, plus the cost of postponing earnings by T years, given by e^{rT}, where r is the annual cost of borrowing, i.e.,

$$G(h) = \beta\{\omega - F - e^{rT}\} \tag{3}$$

So that, combining (2) and (3)

$$G(\omega) = (1 - 1/\sigma)g - (1/\sigma)\beta\{\omega - F - e^{rT}\} \tag{4}$$

and for positive ω the relative wage converges to a value

$$\omega^* = F + e^{rT} + g(\sigma - 1)/\beta \tag{5}$$

From this, we can see that SBTC (and $\sigma > 1$) does not lead to an ever-increasing wage premium. The move to a new, constant, rate of technological progress leads to an increase in wage dispersion but we should not expect this to continue, since supply adjusts.

We have set out this theory for two reasons. The first is that, when looking at the data, we need to distinguish between a continuing upward trend and an upward shift in the degree of wage dispersion. If, empirically, wage dispersion has ceased to increase, this does not mean that SBTC (or globalization) has come to an end.[8] Indeed, from (5), we can see that g falling to zero would imply that the wage premium fell back to its earlier value. The second reason for the explicit model of dynamics is that economic theory is valuable because it points to other mechanisms that may be important. From (5), we can see immediately that the same forces—SBTC or globalization—can have differing effects in different countries depending on the speed of adjustment of supply (via β). This is one response to the challenge to the SBTC explanation made by Lemieux: "if technological change is the explanation for growing inequality, how can it be that other advanced economies subject to the same technological change do not experience an increase in inequality?" (2008, p. 23). A country where the labor market is more responsive will see a smaller increase in wage dispersion. From (5), we can also see that wage dispersion may increase on account of increases in the cost of education. Raising student fees leads, with these market responses, to a higher wage premium. Such a rise in wage dispersion does not however imply a rise in inequality—when viewed over the life time as a whole.

4.2 Steady States and Transitional Dynamics

This condensed presentation of the supply and demand model of wage inequality shows the power of theory. At the same time, it raises questions as to whether the available models actually capture what is being observed and provide a reasonable basis for projecting the future development and drawing policy implications. On the one hand, there is the issue of the relative strength of the various mechanisms simultaneously at play—e.g., biased technical change, educational choices, supply of skills, capital accumulation

[8] The model has been discussed in terms of technological change, but similar considerations apply where demand is shifting on account of globalization.

and allocation, etc. On the other hand, there is the issue of the time scale, which is a key aspect that we would like to highlight—both here and, later, when we discuss the distribution of wealth.

To be tractable, the analysis generally focuses on the steady-state or long-run equilibrium properties of these models, without necessarily mentioning how long the long run is. In Chapter 14, Vincenzo Quadrini and Victor Rios-Rull thus warn the reader that: "for simplicity of exposition, we limit the analysis here to steady-state comparisons with the caveat that in the real economy, the distribution will take a long time to converge to a new steady state." However, to bring theory to bear on the observed evolution of the distribution of income and on policy, tackling transitional phases is essential. In distributional analysis, the fact that accumulation of human and physical capital, the factors the most likely to modify the distribution of income, takes time at both the individual and the aggregate level means that it may take a long time for the economy to adjust to an exogenous shock. After all, if a shock modifies the incentive to get tertiary education, it will take around 40 years—or the duration of active life—before the full effect is felt, or in other words all workers in the labor force have faced the new trade-off between education and work. In between, many other things may change.

To illustrate this point, consider the dynamics of the preceding model in adjusting to a steady state, where again we focus on SBTC. This requires modeling with a little more detail the behavior of h, the ratio of the skilled, L_h, to the unskilled, L_b, labor force. Assume the population is stationary so that a proportion n of the labor force is exiting every year and an equal proportion is entering. For a stationary population, n would be the inverse of the duration of active life, roughly 2.5%. The important point is that most of the change in the skill structure of the labor force goes through its progressive renewal and modifying educational choices made by the entrants. More precisely, assume that the dynamics of the skill structure of the labor force is given by:

$$\Delta L_h = n\left[1 + \beta\left(\omega - F - e^{rT}\right)\right] L_h - nL_h$$

$$\Delta L_b = nL - n\left[1 + \beta\left(\omega - F - e^{rT}\right)\right] L_h - nL_b$$

where L is the total labor force ($L_h + L_b$). In other words, the rate of growth of the skilled labor force depends on the net benefit from acquiring a skill from more schooling, whereas the growth of the unskilled labor force is driven by those entrants who have decided not to stay in school. Dividing these two equations respectively by L_h and L_b and subtracting the latter from the former yields:

$$G(h) = n(1 + h)\beta\left(\omega - F - e^{rT}\right) \tag{3'}$$

which is a slight modification of (3). Then the dynamics of the economy when skill-biased technical change (SBTC) takes place, and the productivity ratio (a_h/a_b) increases at the constant rate g, is given by (3′) and:

$$G(\omega) = (1-(1/\sigma))g - (1/\sigma)G(h) \tag{2}$$

Note that, with this modification of the labor supply dynamics, no steady state exists in the (2)–(3′) model as long as g is strictly positive. Simulating this dynamic system with $\sigma=1.2$, $h(0)=3$, $n=1/42$, $F=1.9$, $r=2.5\%$, $T=4$, $\beta=1$, and with no SBTC ($g=0\%$) yields an initial steady state. Then, in year 1, the rate of SBTC rises to $g=3\%$, and the economy evolves according to the trajectories shown in Figure 4 for the wage skill ratio or skill premium, ω, and for the size of the skilled labor force relatively to the number of unskilled workers, h. As before, the wage skill ratio rises, but—there being no steady state with positive g—it then falls back to the starting level.

Three features are of interest in Figure 4. The first is that, as expected, the wage skill ratio increases and then turns back down as the relative supply of skilled labor increases. Yet, this stabilization takes approximately 30 years to materialize. Even after 50 years, the curve has only just begun to turn down. The second feature is that the overall increase in the skill premium is rather modest, 3.1% in Figure 4. Such a limited increase is due of course to the labor supply response or the value of the elasticity β. Without such a response, the skill premium after 30 years of SBTC at the rate of 3% would have been 12% higher than initially. The third interesting feature in Figure 4 is the behavior of the skilled share of the labor force, which keeps increasing even when the wage skill premium has stabilized. In the presence of SBTC, such a persistent increase in the proportion of

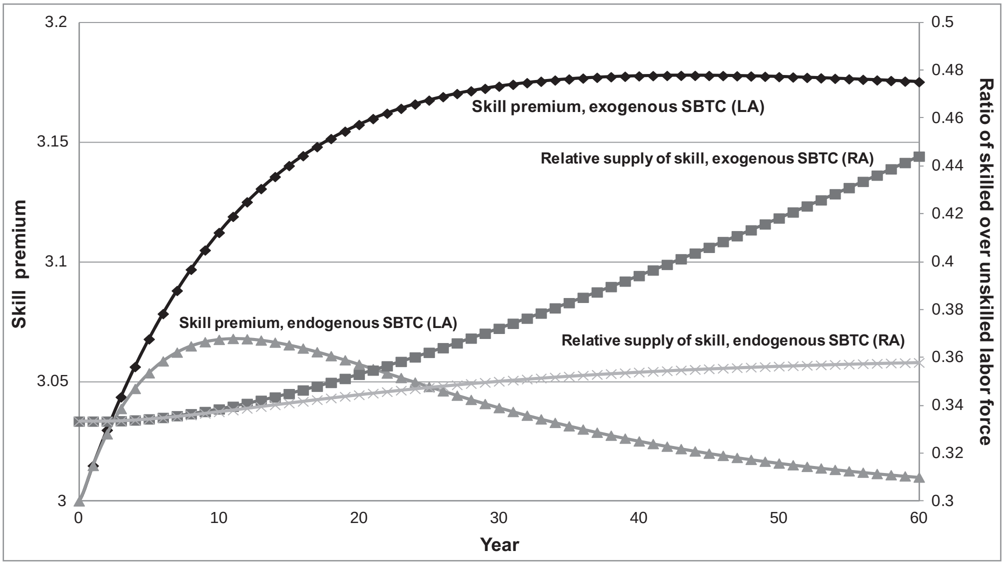


Figure 4 Skill-biased technical change: simulated trajectories of the skill premium and the relative supply of skilled labor with exogenous or endogenous rate of skill-biased technical progress. *Note: The skill premium is shown on the left-hand axis; the relative supply of skilled workers on the right-hand axis.*

skilled workers in the labor force is indeed needed to stabilize the wage skill premium and occurs because the long-run equilibrium premium is above the cost of acquiring a skill relatively to the unskilled wage level, unlike the case at the initial equilibrium.

Economic theory is thus important to understand switches from a long-run equilibrium to another but also the nonsteady-state behavior. In this respect, note that the transition modeled above depends on the way expectations are formed. Equation (3) and (3′) are implicitly based on static expectations about the skill premium. Rational expectations are somewhat irrelevant in the present framework as the cause of the increase in the skill premium is not necessarily known by economic agents. But adaptive expectations might yield another time path. Likewise, a stronger supply response, β, to a change in the skill premium makes the transition shorter, and the new long-run wage skill ratio smaller.

From the point of view of the inequality of earnings, Figure 4 shows that two forces are at play. On the one hand, the increase in the wage differential increases inequality in the sense that the Lorenz curve shifts outward. On the other hand, the fact that more people get skilled has an ambiguous effect on the distribution of earnings. As the average earnings increases, both low-skill and high-skill people lose relatively to the average—the bottom of the Lorenz curve shifts outward but the opposite occurs at the top. The relationship between these two ratios, ω and h, and the overall inequality as summarized by the Gini coefficient is discussed in Chapter 18. Note however, that this analysis refers to the distribution of earnings. Implications of the joint dynamics of technology and education for the inequality of income might be different because of various mechanisms including endogamy, joint labor force participation within couples or fertility differentials.

4.3 Endogenous Technological Change

To this point, technological change has been assumed to be exogenous, but the change in relative wages may induce a change in the degree of bias. In 1932, Sir John Hicks reasoned in his *The Theory of Wages* that "a change in the relative prices of the factors of production is itself a spur to invention, and to invention of a particular kind—directed to economizing the use of a factor that has become relatively expensive" (1932, pp. 124–125). This was later formalized in terms of the bias between capital- and labor-augmenting technological change by Kennedy (1964), Samuelson (1965), and Drandakis and Phelps (1965).[9] More recently, it has been taken up in the form of "directed technical change" by Acemoglu (for example, 2002).

[9] We may note that the induced technical progress literature identified the key role of the elasticity of substitution in determining the stability of the dynamic process governing the degree of bias: the condition for stability of the steady state in the model of Drandakis and Phelps (1965) is that the elasticity of substitution between capital and labor be less than 1. But the wage dispersion literature takes it for granted that the elasticity is greater than 1, which would be relevant if the same models were to be applied to the bias between skilled and unskilled workers.

Where technological progress is endogenous, this may substantially modify the trajectory of the economy. Following Acemoglu (2002), assume that the bias of technical progress is determined by the relative profitability of improving the productivity of the two factors of production. This profitability depends itself on two effects: a price effect, defined as the price of the technical bias—here the ratio $\omega/(a_h/a_b)$—and a market size effect according to which technical progress should favor the factor relatively the most abundant—i.e., the relative supply of skilled labor, h. Assuming that these two effects are multiplicative, and within the same CES framework as in (1), the relative profitability of investing in productivity gains of skilled labor thus depends on (Acemoglu, 2002, p. 790):

$$\frac{h\omega}{a_h/a_b} = A\left(\frac{a_h}{a_b}\right)^{-1/\sigma} h^{1-(1/\sigma)}$$

Then, the dynamics of the productivity differential can be specified as:

$$\Delta\left(\frac{a_h}{a_b}\right) = g.\left(\frac{a_h}{a_b}\right) = \gamma\left[A\left(\frac{a_h}{a_b}\right)^{-1/\sigma} h^{1-(1/\sigma)} - c\right] \tag{6}$$

where c stands for the relative cost of developing technical progress in one factor in comparison with the other and γ is the response rate of actual technical change to the economic incentive.

Adding Equation (6) to the preceding model (2)–(3′) modifies substantially the dynamics of the skill premium and relative labor supply. In the simulation shown in Figure 4, c is chosen so that (6) is initially stationary. Then an exogenous drop in the cost parameter, c, in year 1 triggers SBTC, initially at a rate identical to the exogenous rate, g, in the previous simulation (the response rate, γ, is taken to be 1). If the new trajectory is initially similar to the one obtained before, a divergence occurs after a few years. The rate of growth of the skill premium declines and a turning point is reached after 10 years. Then the skill premium starts falling as SBTC attenuates due to the negative price effect. If the relative supply of skilled labor keeps increasing—because the skill premium keeps being above its initial value—its rate of growth is much smaller than in the previous simulation. Interestingly enough, the new steady state to which the economy is very slowly converging displays the same wage skill premium as initially, but a larger relative supply of skill. In other words, in the very long run the fact that the cost of improving the relative productivity of skilled workers has fallen simply resulted in an increase in relatively more people being skilled. In the short run, however, the skill premium moved up.

This analysis demonstrates the value of theory in understanding the mechanisms behind the evolution of a simple aggregate inequality indicator like the wage skill premium. It shows the need to consider transitional paths between equilibria, not just steady states, as well as the multiplicity of mechanisms influencing specific economic magnitudes to interpret their observed evolution.

4.4 Beyond Supply and Demand

The supply and demand story assumes that all agents act as price-takers: that we have perfect competition. As was observed by Michael Kalecki, "perfect competition—when its real nature, that of a handy model, is forgotten—becomes a dangerous myth" (1971, p. 3). In the real world, there are firms that have market power, as do collective organizations such as trade unions, and market power affects the operation of the labor market. The relative bargaining power of different actors determines the way in which economic rents are shared and hence the distribution of income. Power in turn is affected by the legal rights of workers, their representatives, and of employers. Such considerations turn the spotlight on governments, where the trend of recent decades has been to scale back the rights of workers, but also on employers. Where employers have market power, they can make choices regarding their employment practices, such as, for example, adopting the policy of paying "a living wage" or of limiting the range between top and bottom pay in their enterprise.

Bargaining power is not limited to firms and unions, as is shown by search and matching models of the labor market that involve individual workers and employers. Frictions in the labor market mean that, while *ex ante* competition may drive down the expected value of filling a job vacancy to the cost of its creation, *ex post* the matching of a worker to a vacancy creates a positive surplus. Without a positive surplus, no jobs are created. The worker offered a job has a degree of bargaining power, since if he or she rejects the job offer, the employer has to return to the pool with the risk that no match can be secured. The magnitude of the risk, and hence the worker's leverage, depends on the tightness of the overall labor market; the worker's leverage also depends on the cost of remaining unemployed. The impact on the distribution of earnings depends on how bargaining power varies across jobs, but the important point is that market forces, even in a globalized world, impose only upper and lower limits on differentials. This becomes particularly important when there are multiple possible market outcomes. Atkinson (2008) suggests a behavioral model of changing pay norms, where there is more than one locally stable equilibrium consistent with profit-maximizing behavior by employers. What has been observed in recent decades may be a shift from one equilibrium to another with much wider pay dispersion, particularly at the top.

There is therefore considerable scope for social institutions and social norms to influence the degree of pay dispersion, as is discussed in Chapters 18 and 19. In the latter, Förster and Tóth note that "while it is widely recognized that institutions matter as an important factor for identifying the multiple causes of inequality . . . the weight attached to this factor in econometric studies has for a long time been limited" (p. 1801). As is stressed by Salverda and Checchi in Chapter 18, there is a pressing need to bring together the two rather separate literatures on supply/demand explanations and on institutions.

4.5 Bringing Back Capital

The SBTC explanation of rising wage dispersion focuses on the labor market, but—as the presence of the term e^{rT} in Equation (3) hints—we need to consider, not just the labor market, but also the capital market. Stated in terms of the aggregate production function, we have to consider not just $F(L_b,L_h)$, where L_b denotes basic-educated labor and L_h denotes higher-educated labor, but $F(K,L_b,L_h)$, where K is capital. We have to consider, not just the ratio of skilled to unskilled wages, but also the relative shares of wages and profits—the classical problem of distribution. Indeed, as in the classical analysis, we should extend the production function to include land and natural resources, N, giving output as $F(K,L_b,L_h,N)$.

The extension to three, or more, factors means that substitutability and complementarity becomes more complex, with richer potential outcomes for the distribution of income. The interesting possibilities include capital being a substitute for basic-educated workers but a complement to higher-educated workers. Such models, in the line of Krusell et al. (2000) are among those discussed in Chapter 14 by Quadrini and Rios-Rull. An alternative has been proposed by Summers in his Feldstein Lecture (2013). As Summers notes, capital can now be seen as playing two roles: not only directly via the first argument of the production function but also indirectly insofar as it supplants human labor through robotization. Denoting the first use of capital by K_1 and the second by K_2, the aggregate two-factor production function becomes $F(K_1, AL+BK_2)$, where A and B depend on the level of technology. The production function is such that capital is always employed in the first use, but may or may not be used to supplement labor. The condition under which robots, or other forms of automation, are used to replace human labor depends, as one would expect, on the relative costs of labor and capital. Where there is perfect competition, K_2 is zero where the ratio of the wage, w, to the rate of return, r, is less than A/B, and where K_2 is positive, then $w/r=A/B$. The ratio of the wage share to the capital share in the latter case is $(A/B)/(K/L)$ and falls with the capital–labor ratio.

We can therefore tell a story of macroeconomic development where initially the Solow model applies: the capital stock is below the level at which w/r exceeds A/B. In this context, a rising capital–labor ratio leads to rising wages and a falling rate of return. The capital share rises if and only if the elasticity of substitution between capital and labor is greater than 1 (about which there is debate—see Acemoglu and Robinson, 2014, footnote 12). Beyond a certain point, however, the wage/rate of return ratio reaches A/B, and K_2 begins to be positive. We then see further growth in the economy, as capital per head rises, but the wage/rate of return ratio remains unchanged. There is no longer any gain to wage-earners, since they are increasingly being replaced by robots/automation. What is more, the capital share rises, independently of the elasticity of substitution. It is as though the elasticity of substitution is increased discontinuously to infinity. In this way, the textbook Solow growth model can be modified in a simple way to highlight the central distributional dilemma: that the benefits from growth now increasingly accrue through rising profits. This outcome was indeed stressed some 50 years ago by James Meade in his *Efficiency, equality and the ownership of property* (1964), where he argued with considerable prescience that automation would lead to rising inequality.

4.6 The Distribution of Wealth

The distribution of wealth is the subject of the long-run studies in Chapter 7 by Roine and Waldenström and Chapter 15 by Piketty and Zucman. Both chapters show that the concentration of wealth was very high in the eighteenth to nineteenth centuries up until the First World War, dropped during the twentieth century, but has been rising again in the late twentieth and early twenty-first centuries. Chapter 15 shows that in France, the United Kingdom, and other countries, there has been a return of inheritance.

In Chapter 15, Piketty and Zucman begin by saying that

> *to properly analyze the concentration of wealth and its implications, it is critical to study top wealth shares jointly with the macroeconomic wealth/income and inheritance/wealth ratios. In so doing, this chapter attempts to build bridges between income distribution and macroeconomics.*

Building such a bridge is indeed one of our aims in this Introduction, and, with this in mind, we return to the question of the timescale for the macroeconomic magnitudes. Denoting aggregate capital by w_t, aggregate income by y_t, and their ratio by β_t, we have:

$$\beta_{t+1} = \frac{\beta_t + s_t}{1 + g_t} \tag{7}$$

where s_t and g_t are respectively the net saving rate and the growth rate of income at time t. Assuming that those two rates are constant, the steady-state equilibrium of the economy is given by $\beta^* = s/g$. With s equal to 10% and g equal to 3% per annum, the equilibrium capital–income ratio is 3.33. But, then, suppose that growth decelerates and the economy's constant growth rate falls to 2%. The economy will then converge toward a new equilibrium with a capital–income ratio equal now to 5. How long though will it take to get to this new equilibrium? As a matter of fact, the process described by (7) is quite slow. A simple simulation shows that, in going from 3.33 to 5, it will take almost 30 years for the capital–income ratio to reach 4, the double to reach 4.5 and more than a century to reach 4.8. As was shown many years ago by Ryuzo Sato (1963), the adjustment time in the neoclassical growth model can be extremely long.[10] With such a transitional phase, relying on the steady-state properties of a theoretical model may be misleading for economic or policy analysis, even with a horizon of a decade. The direction of changes expected at equilibrium

[10] As was pointed out by Sato (1966), the conclusions reached regarding convergence times are sensitive to the precise assumptions made concerning savings and technical change; the key issue is that transition path should be examined.

because of some exogenous modification or some policy changes is most likely to be felt along the whole transitional trajectory but their size might have to be substantially scaled down at the beginning of the transition. A tax on capital that reduces the saving rate by 1 percentage point, from 10% to 9%, for example, would lead to a drop in the steady-state capital–income ratio of 10%, but only 2.3% after 10 years and 4% after 20 years.

Let us now turn to the *distribution* of wealth. In that case, focusing on steady-states—or the golden rule—leads in some sense to the simple dismissal of distributional issues. Chapter 14 shows that any distribution of wealth combined with any distribution of work abilities is consistent with the steady-state equilibrium of the neoclassical model with dynastic agents, provided that aggregate wealth and aggregate effective labor satisfy some consistency relationship that involves the (common) rate of time preference of agents—the same kind of result holds trivially in endogenous growth models of the AK type, see Bertola et al. (2006, chapter 3). This is fine but possibly of limited practical relevance. Assume indeed that an economy initially at a steady state with a distribution D of wealth is then subject to some shock, for example, a technological shock or an income tax, that modifies its aggregate long-run equilibrium. Then, in moving toward this new equilibrium, the distribution D will change and, at the new equilibrium, there will be a new distribution D'. The fact that this new aggregate equilibrium may be supported by another distribution than D' is not the relevant point. What we are interested in is the change from D to D' and this is certainly not indeterminate. Likewise, the indeterminacy of distribution in a steady state does not mean that redistribution has no macroeconomic effect and no impact on the primary distribution of income. As long as redistribution cannot be lump sum, it will modify both the steady-state equilibrium and the distribution of primary and disposable income.

The models in Chapter 15 of the distribution of wealth are rather different and have different long-run properties. The treatment of time is again important. To be precise, let us take the unit of time as the lifetime (identical for all), with the present value of inherited wealth of individual i denoted by w_{it}. It is assumed that lifetime savings are a constant proportion of the aggregate of wealth and income:

$$w_{it+1} = S_{it}(w_{it} + y_{it}), \text{ with } y_{it} = y_{Lt} + Rw_{it} \tag{8}$$

where y_{Lt} is the lifetime labor income, assumed to be identical across individuals, R is the rate of return over the lifetime, and S_{it}, the individual saving rate defined on wealth plus lifetime income. S_{it} is assumed to be independently and identically randomly distributed around some mean value, S, across periods. Aggregating the accumulation equation over all individuals in a generation yields:

$$w_{t+1} = S(w_t + y_t), \text{ with } y_t = y_{Lt} + Rw_t \tag{9}$$

Combining (8) and (9) and assuming that the aggregate economy has converged to a steady state, it is shown in Chapter 15 that the dynamic behavior of the wealth of individual i, relatively to the mean wealth of the population, z_{it}, is given by the following multiplicative stochastic difference equation:

$$z_{it+1} = \frac{S_{it}}{S}[(1-\varphi+\varphi z_{it}], \text{ with } \varphi = S\frac{1+R}{1+G} \tag{10}$$

where G is the rate of growth over a lifetime.[11] Under the condition that $\varphi < 1$, the steady-state stochastic distribution of z_{it} has a Pareto upper tail, with a Pareto coefficient that decreases with φ, where a smaller coefficient corresponds to greater concentration of wealth. Denoting the annual rate of interest by r and the annual rate of growth by g, and assuming a lifetime of H years, φ can be expressed as $Se^{(r-g)H}$. It follows that long-run wealth concentration increases with $r-g$; it is also clear that concentration increases with the savings rate, S. Both elements have a role to play.

Suppose, instead, that one adopts an *intragenerational* perspective, with the time unit as the year rather than a lifetime, and assumes that people save a proportion of their current income, s_{it}, drawn randomly from some distribution with expected value s. Then (8) and (9) are transformed into:

$$w_{it+1} = w_{it} + s_{it}(y_t + rw_{it}); w_{t+1} = w_t + s(y_t + rw_t) \tag{11}$$

where y_t is now the common annual wage income. Assuming a steady state with growth rate, g, and using the same kind of derivation as above, the stochastic difference equation (10) becomes[12]

$$z_{it+1} = z_{it}\left[\frac{1}{1+g} + \frac{s_{it}}{s}\frac{rs}{1+g}\right] + \frac{s_{it}}{s}\frac{g-rs}{1+g} \tag{12}$$

Under the assumption that $E(rs/(1+g))(s_{it}/s) + (1/(1+g)) < 1$ or $rs < g$, the distribution of z_{it} converges toward a steady-state distribution with a Pareto upper tail, with a

[11] The derivation of that equation goes as follows. At a steady state: $w_{t+1} = (1+G)w_t$. From (8) and (9), it then follows that $z_{it+1} = (S_{it}/S)S((1+R)/(1+G))z_{it} + S_{it}(y_{Lt}/w_{t+1})$. But (9) implies at a steady state that: $w_{t+1} = S[y_{Lt} + (1+R)\cdot(w_{t+1}/1+G)]$. Then: $S(y_{Lt}/w_{t+1}) = 1 - S((1+R)/(1+G))$ and (10) follows.

[12] As before, at the steady state, $w_{t+1} = (1+g)w_t$. From (11), we obtain:

$$z_{it+1} = \frac{z_{it}}{1+g} + \frac{s_{it}}{s}\frac{rs}{1+g}z_{it} + \frac{s_{it}y_t}{w_{t+1}}. \tag{13}$$

Then, the second part of (11) becomes:

$$w_{t+1} = \frac{w_{t+1}}{1+g} + sr\frac{w_{t+1}}{1+g} + sy_t.$$

This implies:

$$\frac{y_t}{w_{t+1}} = \frac{g-rs}{s(1+g)}$$

and then plugging that expression back into (13) leads to (12).

Pareto coefficient that decreases, and a wealth concentration that increases, with $rs/(1+g)$. This refers to the distribution of current wealth, since there is a fresh drawing of s_{it} each period of life. (The independence assumption has therefore quite different implications.) In this model, it is the balance between rs and g that determines the long-run distribution, as in the early models of Meade (1964) and in the primogeniture version of Stiglitz (1969).[13]

Which model is the most appropriate? With the long horizon adopted to study the evolution of the distribution of income and wealth in Chapters 7 and 15, it would seem that the intergenerational framework is the most appropriate. It can also be argued that the assumption about randomness, with persistent lifetime good or bad luck, captures better our distributional concerns. Against this, it is the distribution of current wealth that is observed (as, for example, in Figure 1). It has, however, been shown by Benhabib, Bisin, and Zhu that when a model of lifetime wealth accumulation, with a realization fixed for any household during its lifetime, is embedded in a model of the current distribution of wealth, then "the power tail of the stationary distribution of wealth in the population is as thick as the thickest tail across the wealth distribution by age" (2011, p. 132). Put loosely, the upper tail of the current distribution tends to be dominated by the most unequal generation. But, it remains the case that the full effect of an increase in r–g from today onward is bound to be observed only some generations from now. As a matter of fact, the inegalitarian effect that will be observed in the second or even third generation might be very limited and r–g may change again in the distant future.

The conclusions that we draw are twofold. The first is that, as in the discussion of mobility in Chapter 10, it is necessary to consider both the intra- and the intergenerational dimensions, and that in both cases a better understanding of the transitional periods seems crucial. The second is that the evolution of the distribution of wealth depends on savings behavior, on the rate of return, and on the rate of growth. In this context, we should not forget that there were two arms to Kuznets (1955) Presidential Address, in which he sought to explain why inequality was at that time falling despite the existence of long-term forces leading to higher inequality. One arm was the theory of structural change that has come to characterize his approach, but the other was the concentration of savings in the upper income groups. This led him to conclude that "the basic factor militating against the rise in upper income shares that would be produced by the cumulative effects of concentration of savings, is the dynamism of a growing and free economic society" (1955, p. 11).

[13] It is rs–g that formed the basis for the time-series analysis of the share of the top 1% of wealth-holding in Great Britain over the period 1923–1972 carried out by Atkinson and Harrison (1978, chapter 9). They showed the significance of two variables likely to influence the rate of accumulation (rs): the share price index and the rate of estate taxation.

5. THE ROLE OF POLICY

The role of policy is considered in many of the chapters, but this is the explicit focus of the final part of the Handbook.

5.1 Policy Objectives

Here, we should begin by observing that the past 15 years has seen a major change in the extent to which there has been official adoption of distributional objectives. This development is the culmination of a series of shifts in attitudes toward policy, notably with regard to the abolition of poverty. In Chapter 22, Martin Ravallion traces, with a broad geographical and historical sweep, the evolution of thinking on poverty and antipoverty policy.

The most evident manifestation of this change has been the adoption at a world level of the Millennium Development Goals (MDGs). The goals were ratified by world leaders in 2000 at the U.N. Millennium Assembly, and the first on the list was the halving between 1990 and 2015 of the proportion of people whose income was less than \$1 a day (later \$1.25 a day). At a national level, countries have adopted their own goals, such as the national social target for poverty reduction in Ireland aiming to reduce consistent poverty. In the United Kingdom, the Child Poverty Act 2010 requires the government to produce a poverty strategy every 3 years setting out actions to end child poverty. At a regional level, the European Union adopted in 2010, as part of the Europe 2020 program, the objective of reducing by 20 million the number of people in or at risk of poverty and social exclusion. These have been translated to varying degrees into national targets (Social Protection Committee, 2014).

It is not clear whether the same kind of change is occurring in the field of inequality reducing policies. Due to the rise in inequality, and possibly to the recent crisis, it is certainly the case that the public spotlight is focused on inequality and some announcements have been made by politicians that important measures would be taken to fight inequality. Yet, few explicit targets have been set and no ambitious measure has been taken or is being seriously considered in advanced countries that would make a major dent in the existing income inequality.

5.2 Impact of Policy to Date

What grounds are there then for optimism if one is concerned with present levels of economic inequality? Can we point to past experience where inequality has been reduced? The first obvious, but important, point is that inequality is not increasing everywhere. Globally, the recent past has been more encouraging, as is well summarized in Chapter 9 by Alvaredo and Gasparini:

> *The available evidence suggests that on average the levels of national income inequality in the developing world increased in the 1980s and 1990s, and declined in the 2000s. There was a remarkable fall in income poverty since the early 1980s, driven by the exceptional performance*

> *of China over the whole period, and the generalized improvement in living standards in all the regions of the developing world in the 2000s.*

They caution that the decline in the 2000s was not universal: in 15% the fall was less than 2.5 percentage points, and in 20% of cases the Gini coefficient increased. The latter included two populous countries: China and Indonesia. The decline was most evident in Latin America, where they say

> *This remarkable decline appears to be driven by a large set of factors, including the improved macroeconomic conditions that fostered employment, the petering out of the unequalizing effects of the reforms in the 1990s, the expansion of coverage in basic education, stronger labour institutions, the recovery of some countries from severe unequalizing crises and a more progressive allocation of government spending, in particular monetary transfers.*

In other words, policy was relevant in influencing both market incomes and redistribution. In the case of Brazil, they conclude that the two major determinants were the decline in wage inequality, due to the expansion of the supply of educated workers and to the substantial increase in the minimum wage, and the expansion of cash transfers, notably the *Bolsa Família*. Despite this noticeable progress, however, it remains the case that Brazil, like most other countries in Latin America, is extremely unequal by world standards—with the exception of several countries in sub-Saharan Africa. The drop in inequality observed during the 2000s compensated for the increase that took place during the 1980s and part of the 1990s. Over the very long run, progress remains limited. Moreover, it must be kept in mind that top income earners are undersampled in the developing countries' household surveys so that it is not impossible that reported inequality figures miss the same increase at the top of the distribution as has been observed in many developed countries.[14]

Although lower than in Latin America, inequality is sizable in many emerging Asian countries and it is increasing. In discussing how far the policies pursued in several Latin American countries could be applied in an Asian context, Kanbur notes in Chapter 20 that

> *the additional expenditure on conditional cash transfers requires revenues, and the progressivity of the tax system is another major determinant of how globalization related increases in inequality can be mitigated. But progressivity is also important in addressing the rise in very high incomes the world over, especially in Asia. Asian tax systems do not generally score highly on progressivity. In fact, it is argued that raising progressivity of taxation would have a bigger impact on inequality than elsewhere in the world.*

[14] Alvaredo and Gasparini warn the reader in Chapter 9 that in the household surveys on which they are drawing there is substantial understatement of the top incomes. Cornia notes, in his analysis of recent distributive changes in Latin America, that "given the scarcity of information on capital incomes and the income of the 'working rich' in household surveys [it is not possible] to establish formally whether the distributive changes . . . concern also the top percentiles of the income distribution" (2014, p. 7).

What about the richer countries? In terms of reducing the inequality of market incomes, the standard policy response is educational expansion, as would be indicated by the supply and demand explanation for rising wage dispersion (as discussed in Section 4). The review of cross-country time series evidence by Förster and Tóth in Chapter 19 concludes that most evidence points to an equalizing impact of educational expansion:

> *none of the studies covering the set of OECD/EU countries suggest a dis-equalising role of the growth in average educational attainment over the past three decades, but to the contrary, in their majority rather an equalising one. Human capital can be seen as a complement to technology. Increases in human capital and in the supply of skills are necessary to decrease and eventually reverse the pressure to higher inequality that stems from technological change.*

The impact of labor market policy is reviewed in Chapter 18 by Salverda and Checchi who conclude that their empirical results

> *are consistent with the main findings in the literature. … They confirm that the presence and stringency of a minimum wage reduces earnings inequality, also setting an (implicit) control on the distribution of working hours, which seems to be the main channel of inequality reduction of the bargaining activity of unions. Less common in the literature is the finding of a negative impact of both active and passive labour market policies.*

The reason that the overall distributive effect of labor market regulations and institutions can be insignificant, as found by Förster and Tóth in their cross-country analysis, is that the employment and wage dispersion effects can operate in opposite directions.

When it comes to redistributive tax and transfer policy, Förster and Tóth conclude in Chapter 19 that:

- Policies are inequality reducing, but the effects vary across countries;
- Transfers are typically more effective than taxation;
- There has been a reduction in the redistributive effectiveness since the 1990s;
- Behavioral responses may offset but do not in general outweigh the first-round effects.

The last of these conclusions is particularly important. For understandable reasons, much of the analysis of public policy by economists in recent decades has focused on negative behavioral responses. Understandable, since the toolkit of economists is designed to illuminate these responses and the second-round effects are often missed in the public debate. At the same time, the analysis seems often to lose sight of the purpose for which transfers are paid. As put by Ive Marx, Brian Nolan and Javier Olivera in Chapter 23, "no advanced economy achieved a low level of inequality and/or relative income poverty with a low level of social spending, regardless of how well that country performed on other dimensions that matter for poverty, notably employment." The GINI project concluded that

> *The best performers among the rich countries in terms of employment and economic and social cohesion have one thing in common: a large welfare state that invests in people, stimulating them and supporting them to be active and adequately protecting them when everything else fails. This continues to offer the best prospect for rich countries pursuing growth with equity.*
>
> **Salverda et al. (2014, p. 349).**

But it is also clear, as these authors bring out, that it is not only the aggregate but also the design of spending that matters. It is for this reason that the construction of policy reforms has to be based on analysis of the contribution that they can make to distributive and efficiency objectives, and to this end an important development has been that of microsimulation modeling, as surveyed in Chapter 24 by Francesco Figari, Alari Paulus and Holly Sutherland. These models build on the improvements in data availability described in Section 1—their construction requires access to microdata. They also require in-depth knowledge of the institutional details of public policy and how it operates in reality. As is discussed in Chapter 24, there is a considerable challenge in modeling noncompliance and—the less commonly discussed—other side of the coin, which is the non-take-up of benefits to which people are entitled. In melding microdata and institutional detail, the microsimulation models provide an important bridge between the theoretical analysis of policy design and the implementation of policy in the form of legislation and administration.

5.3 Prospects for Future Policy

Given their high level of inequality and the limited development of redistribution in many developing countries, the scope for progress in redistributive policies and in pre-redistribution opportunity equalizing policies is considerable. This is especially true in middle income or emerging countries where state capacity is sufficient to manage effective redistribution instruments.

What should make things easier for emerging countries' governments wishing to strengthen their redistribution system is, on the one hand, that they can learn from the experience of developed countries and, on the other hand, that modern technology permits better monitoring and control of individual incomes. Beyond some level of income, it is difficult today to function without a bank account and a credit/debit card, so that individual transactions are recorded. Approaching 50% of all households in Latin America hold a bank account and this proportion is rising so that the tax authority's capacity of auditing taxpayers suspected of underreporting their income is necessarily increasing. Yet, the income tax is severely underdeveloped in most emerging countries, where it represents most often less than 2% of GDP. Brazil is an exception but with an income tax amounting to 6% of GDP, it is still lower than the 9% average rate observed in OECD countries. Modern technology also makes easier the transfer of income to people at the bottom of the distribution. Smart payment cards in particular assist in avoiding suspicious leakages. There is therefore scope for extending redistribution and making use of policies in the field of education, social protection, minimum pensions, or minimum wage that are currently seldom used or at a very low scale despite their huge equalizing potential in most countries. As a result, the extent of redistribution in emerging countries, evaluated with the same microsimulation tools as described in Chapter 24, appears much smaller than in developed countries. For instance, the fall in the Gini coefficient when moving from market income—including replacement income like public or private pensions—to disposable income averages 3 percentage points in Latin American countries where such simulation has been performed (Lustig, 2014), whereas it oscillates around 10 percentage points for rich countries (Immervoll et al., 2009).

Of course, having the capacity to redistribute will lead to a substantial redistribution only if there is the political willingness to do so, or if the political system permits a majority to impose some redistribution, as could be expected in a democracy. In the empirical work undertaken in Chapter 21, Acemoglu et al. find that a transition to democracy indeed tends to raise the average tax rate in a country. Yet, no significant effect is found on income inequality as such. This may suggest that the political system is more complex than the mere distinction between democracies and nondemocracies would imply; or, as the authors note, may reflect the poorer quality of the income inequality data, an aspect we discussed in Section 3.

In contrast, OECD countries may be closer to their frontier in terms of the trade-off between less inequality and a higher degree of aggregate economic efficiency. The distance depends however on the institutional features of taxes and transfers—aspects to which economists have devoted too little attention (Atkinson, 1999)—and there may be scope for new and innovative ideas, as discussed in the final part of this section. The frontier itself may have been affected itself by the globalization process, its requirement for more competitiveness and, through factor mobility, its weakening of redistributive instruments like the progressivity of income taxation, including capital or capital income. At the same time, the same forces of globalization may have increased the need for social protection, just as they did in the early days of the welfare state toward the end of the nineteenth century. Today, as then, the extent to which redistribution takes place depends on the political context. It may also be the case that the perception of inequality in the society today may not coincide with the evolution of inequality as measured by statisticians and economists. In the United States, for instance, the feeling that income mobility matters more than income inequality and the (unfounded) belief that income mobility is and remains high in comparison with other countries may make the public opinion insensitive to the mounting objective income and wealth inequality. Clearly, this cannot continue forever. At some stage, beliefs will change and it is not unlikely that this process is under way—as McCall (2013) seems to detect in the changing discourse on inequality in the U.S. media.

5.4 Thinking Outside the Box

At the beginning of this Introduction, we evoked the increasing attention being paid by policymakers to rising income inequality. To date, the response in terms of policy proposals has been along conventional lines, notably investment in education and reform of redistribution. In our view, these are important, but if progress is to be made, then we

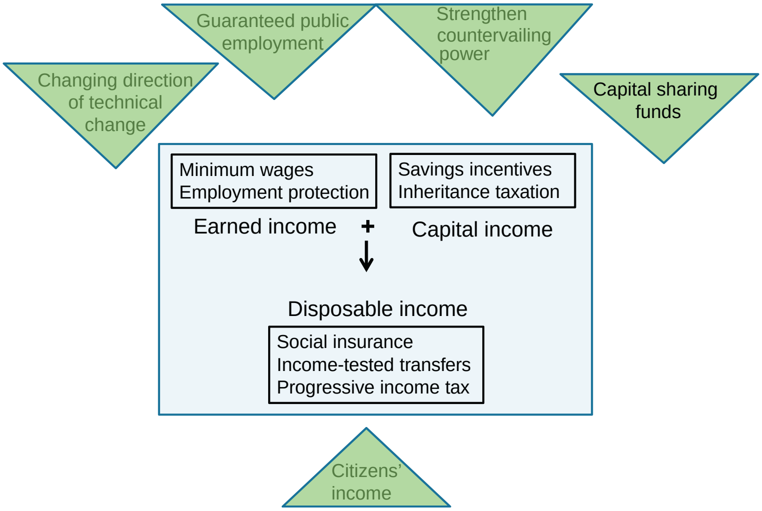


Figure 5 Thinking outside the box.

need to think outside of the box. We must consider ideas that—while far from new—are not on the current policy agenda.

The standard labor market policies—shown inside the box in Figure 5—clearly have a major role to play, and there have been moves to strengthen minimum wages, in order to reduce wage dispersion. As we have noted, however, central is the conjunction of wages and employment, and the latter has proved a hard nut to crack. In our view, one of the crucial elements is the direction of technical progress. Rather than concentrating on factor-augmentation, as in much of the literature, our discussion in Section 4.3 suggests that we should instead focus on the interaction between labor and capital, and specifically the supplantation of labor by capital. Given that much of the innovation is funded directly by public bodies, or subsidized through tax or other concessions, it would be possible to influence this trade-off. This is the first of the "out of the box" alternatives shown in Figure 5. The second concerns public employment. The fact that the present-day market economy does not deliver full employment suggests that we should learn from policy responses in other cases of market failure, notably in financial markets. Here, the government has intervened as a lender of last resort, and the obvious parallel is that the government should act as the employer of last resort. The state should guarantee to everyone seeking it employment at the minimum wage. Such a proposal may seem to some readers outlandish and infeasible on fiscal grounds, but to others it may appear no more outlandish or fiscally irresponsible than the policy that financial institutions are too big to fail. After all, such policies have already been pursued. It is an initiative of this type which was taken by the Indian government when it launched the Mahatma Gandhi National Rural Employment Guarantee Act in 2005. Public employment has formed part of active labor market programs in a number of countries. In the United States, it was authorized under the Humphrey–Hawkins Full Employment and Balanced Growth Act of 1978, which allowed the Federal Government to create a "reservoir of public employment," where these jobs were required to be in the lower ranges of skill and pay to minimize competition with the private sector.

The third proposal refers back to the discussion of bargaining power in Section 4.4. To the extent that rising inequality is the outcome of a shift in the balance of market power favoring profits and capital, its impact may be offset by strengthening the countervailing powers. This may take the form of a stronger role for the social partners, or it may involve more determined action to protect consumers against monopolistic pricing. Again such actions may be rejected as too radical, but again they are not so far removed from current policy. In the case of the European Union, both the promotion of competition and the encouragement of the social partners are already accepted objectives.

The fourth proposal draws attention to an aspect that has been missing from the policy arena—the capital market—but which has received increasing attention following the debate surrounding Piketty (2014). For macroeconomic reasons outlined in Section 4.6, and given the return of inherited capital in a number of richer countries, capital income is potentially of increasing significance, as it was in the past. The particular proposal is indeed far from new. It goes back at least to the eighteenth century Englishman/Frenchman/American Thomas Paine, who in 1797 in his *Agrarian Justice* proposed:[15]

> *To create a national fund, out of which there shall be paid to every person, when arrived at the age of twenty-one years, the sum of fifteen pounds sterling, as a compensation in part, for the loss of his or her natural inheritance, by the introduction of the system of landed property.*

The proposal of Paine for a capital element payable on reaching the age of majority has its modern counterpart in various schemes for asset-based egalitarianism (as proposed, for example, by Ackerman and Alstott, 1999). The creation of a sovereign wealth fund, as already well established in a number of countries, would offer the possibility of a minimum inheritance for all.

The fifth proposal—which also has shades of Paine—is for a citizen's income, or the payment of a guaranteed minimum income to all individuals. Such an income is sometimes described as "unconditional"; however, conditions are naturally attached. As described in Figure 5, a condition would be that of citizenship. An alternative, advocated in Atkinson (1995, 1996) is that of a *participation income*, paid not on the basis of citizenship but of participation in the society in question, through employment, past employment (on retirement), caring for dependants, being available for work when unemployed, in

[15] The text can be downloaded from the Official Web site of the U.S. Social Security Administration. The Web site carries the caution: "this is an archival or historical document and may not reflect current policies or procedures."

approved education or training, and with appropriate provisions for those who are sick, injured, or disabled. The participation income would represent a radical departure from the targeted income-tested transfers that have been the preoccupation of policymakers in recent decades. It would be individual-based, rather than involving a family means-test. It would recognize the fluidity of the employment relationship in the twenty-first century labor market.

No doubt there will be many objections to these final policy proposals, but we hope that the chapters contained in volume II of the Handbook will stimulate new ideas in this important field.

ACKNOWLEDGMENTS

In writing this Introduction, we have drawn heavily on the Handbook chapters. This will be evident from the frequency of cross-references, although we have tried to stop short of the point at which such references become tedious to the reader. We are most grateful to Rolf Aaberge, Facundo Alvaredo, Andrew Berg, Andrea Brandolini, Daniele Checchi, Pierre-André Chiappori, Andrew Clark, Koen Decancq, Jean-Yves Duclos, Francesco Figari, Michael Förster, Marc Fleurbaey, Stephen Jenkins, Salvatore Morelli, Dominique Meurs, Brian Nolan, Jonathan Ostry, Alari Paulus, Sophie Ponthieux, Thomas Piketty, Victor Rios-Rull, Wiemer Salverda, Tim Smeeding, Frederick Solt, Holly Sutherland, István Tóth, Alain Trannoy, and Daniel Waldenström for their very helpful comments on the draft of the Introduction, but none of them should be held responsible in any way for its contents. We thank Maarit Kivilo for her help with the bibliography. Atkinson's research for the Introduction was carried out as part of the EMoD program supported by INET at the Oxford Martin School.

REFERENCES

Acemoglu, D., 2002. Directed technical change. Rev. Econ. Stud. 69 (4), 781–809.

Acemoglu, D., Robinson, J.A., 2014. The Rise and Fall of General Laws of Capitalism. working paper, Cambridge, MA. http://economics.mit.edu/files/9834.

Ackerman, B., Alstott, A., 1999. The Stakeholder Society. Yale University Press, New Haven.

Alkire, S., Foster, J., 2010. Designing the Inequality-Adjusted Human Development Index (HDI): OPHI Working Paper 37. University of Oxford.

Alkire, S., Foster, J., 2011. Counting and multidimensional poverty measurement. J. Public Econ. 95 (7–8), 476–487.

Allison, R.A., Foster, J., 2004. Measuring health inequality using qualitative data. J. Health Econ. 23 (3), 505–524.

Alvaredo, F., Londoño Velez, J., 2013. High Incomes and Personal Taxation in a Developing Economy: Colombia 1993-2010: Center for Equity, Working Paper No. 12. Tulane University.

Anand, P., Santos, C., Smith, R., 2007. The Measurement of Capabilities: Discussion Paper 67. The Open University.

Atkinson, A.B., 1970. On the measurement of inequality. J. Econ. Theory 2, 244–263.

Atkinson, A.B., 1995. Beveridge, the national minimum and its future in a European context. In: Atkinson, A.B. (Ed.), Incomes and the Welfare State. Cambridge University Press, Cambridge, pp. 290–304.

Atkinson, A.B., 1996. The case for a participation income. Polit. Q. 67 (1), 67–70.

Atkinson, A.B., 1999. The Economic Consequences of Rolling Back the Welfare State. MIT Press, Cambridge, MA.

Atkinson, A.B., 2008. The Changing Distribution of Earnings in OECD Countries. Oxford University Press, Oxford.
Atkinson, A.B., Bourguignon, F., 1982. The comparison of multi-dimensioned distributions of economic status. Rev. Econ. Stud. 49 (2), 183–201.
Atkinson, A.B., Bourguignon, F., 2000. Income distribution and economics. In: Atkinson, A.B., Bourguignon, F. (Eds.), Handbook of Income Distribution, vol. 1. Elsevier, Amsterdam.
Atkinson, A.B., Brandolini, A., 2001. Promise and pitfalls in the use of "secondary" data-sets: income inequality in OECD countries as a case study. J. Econ. Lit. 34 (3), 771–799.
Atkinson, A.B., Harrison, A.J., 1978. Distribution of Personal Wealth in Britain. Cambridge University Press, Cambridge.
Atkinson, A.B., Marlier, E., 2010. Indicators of poverty and social exclusion in a global context. J. Policy Anal. Manage. 29 (2), 285–304.
Atkinson, A.B., Morelli, S., 2014. Chartbook of Economic Inequality: Economic Inequality Over the Long Run. INET Oxford. http://www.chartbookofeconomicinequality.com/.
Atkinson, A.B., Casarico, A., Voitchovsky, S., forthcoming. Top incomes and the glass ceiling. Working Paper, INET Oxford.
Becchetti, L., Massari, R., Naticchioni, P., 2011. The Drivers of Happiness Inequality: Suggestions for Promoting Social Cohesion: Working Papers 2011-06. Universita' di Cassino.
Benhabib, J., Bisin, A., Zhu, S., 2011. The distribution of wealth and fiscal policy in economies with finitely lived agents. Econometrica 79 (1), 123–157.
Bertola, G., Foellmi, R., Zweimuller, J., 2006. Income Distribution in Macroeconomic Models. Princeton University Press, Princeton.
Bourguignon, F., 2014. Appraising income inequality databases in Latin America. mimeo, J. Econ. Inequal. (forthcoming).
Bourguignon, F., 2013. La mondialisation de l'inégalité. Le Seuil, Paris (Expanded English version, The Globalization of Inequality, forthcoming, Princeton University Press).
Brunori, P., Ferreira, F., Peragine, V., 2013. Inequality of Opportunity, Income Inequality and Economic Mobility: Some International Comparisons: IZA Discussion Paper 7155. Institute for the Study of Labor (IZA).
Budd, E.C., 1970. Postwar changes in the size distribution of income in the US. Am. Econ. Rev. Pap. Proc. 60, 247–260.
Canada, Statistics, 2013. High-Income Trends Among Canadian Taxfilers, 1982 to 2010, Release January 28, 2013.
Clark, A., Flèche, S., Senik, C., 2012. The Great Happiness Moderation: IZA Discussion Paper 6761. Institute for the Study of Labor (IZA).
Cornia, G.A. (Ed.), 2014. Falling Inequality in Latin America: Policy Changes and Lessons. Oxford University Press, Oxford.
Cornia, A., Rosignoli, S., Tiberti, L., 2008. Globalization and Health Impact Pathways and Recent Evidence: Research Paper No. 2008/74. UNU-WIDER, Helsinki.
Dalton, H., 1920. The measurement of the inequality of incomes. Econ. J. 30 (119), 348–461.
Decancq, K., Fleurbaey, M., Maniquet, F., 2014. Multidimensional Poverty Measurement with Individual Preferences: Princeton University William S. Dietrich II Economic Theory Center Research Paper 058. Princeton University.
Deininger, K., Squire, L., 1996. A new data set measuring income inequality. World Bank Econ. Rev. 10 (3), 565–591.
Drandakis, E., Phelps, E.S., 1965. A model of induced invention, growth and distribution. Econ. J. 76 (304), 823–840.
Duclos, J.-Y., Echevin, D., 2011. Health and income: a robust comparison of Canada and the US. J. Health Econ. 30, 293–302.
Duclos, J.-Y., Sahn, D.E., Younger, S.D., 2011. Partial multidimensional inequality orderings. J. Public Econ. 98, 225–238.
Dutta, I., Foster, J., 2013. Inequality of happiness in the US: 1972–2010. Rev. Income Wealth 59 (3), 393–415.

Ferreira, F., Lustig, N. (Eds.), 2015. International databases on inequality. Special issue of the J. Econ. Inequal., forthcoming.
Fisher, G., 1986. Estimates of the Poverty Population Under the Current Official Definition for Years Before 1959. Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services (Mimeograph).
Fleurbaey, M., Blanchet, D., 2013. Beyond GDP: Measuring Welfare and Assessing Sustainability. Oxford University Press, Oxford.
Foster, J., Lopez-Calva, L., Szekely, M., 2005. Measuring the distribution of human development: methodology and an application to Mexico. J. Hum. Dev. 6 (1), 5–29.
Hicks, J.R., 1932. The Theory of Wages. Macmillan, London.
Immervoll, H., Levy, H., Nogueira, J.R., O'Donoghue, D., Bezera de Siqueira, R., 2009. The impact of Brazil's tax and benefit system on inequality and poverty. In: Klasen, S., Nowak-Lehmann, F. (Eds.), Poverty, Inequality and Policy in Latin America. In: CES-Ifo Seminar Series, The MIT Press, Cambridge, MA, pp. 271–302.
Jenkins, S., 2014. World Income Inequality Database: an assessment, mimeo. Final version forthcoming in J. Econ. Inequal.
Kalecki, M., 1971. Selected Essays on the Dynamics of the Capitalist Economy 1933-1970. Cambridge University Press, Cambridge.
Karoly, L., 1992. The Trend in Inequality Among Families, Individuals, and Workers in the United States: A Twenty-Five-Year Perspective. Rand, Santa Monica.
Kennedy, C., 1964. Induced bias in innovation and the theory of distribution. Econ. J. 74 (295), 541–547.
Kolm, S.-Ch, 1966. The optimal production of justice. In: Guitton, H., Margolis, J. (Eds.), Public Economics. Proceedings of an IEA conference held in Biarritz, published (1969). Macmillan, London, pp. 145–200.
Kolm, S.-Ch., 1971. Justice et équité. Cepremap, Paris (English translation, Justice and Equity, MIT Press, 1988).
Kolm, S.-Ch., 1977. Multi-dimensional Egalitarianism. Q. J. Econ. 91 (1), 1–13.
Kopczuk, W., Saez, E., 2004. Top wealth shares in the US, 1916-2000: evidence from the Estate Tax returns. Natl. Tax J. 57, 445–487 (longer version in NBER Working Paper 10399).
Korinek, A., Mistiaen, J., Ravallion, M., 2006. Survey nonresponse and the distribution of income. J. Econ. Inequal. 4 (1), 33–55.
Krusell, P., Ohanian, L., Rios-Rull, V., Violante, G., 2000. Capital skill complementarity and inequality: a macroeconomic analysis. Econometrica 68 (5), 1029–1053.
Kuznets, S., 1955. Economic growth and income inequality. Am. Econ. Rev. 45 (1), 1–28.
Lemieux, T., 2008. The changing nature of wage inequality. J. Popul. Econ. 21 (1), 21–48.
Lustig, N., 2014. Taxes, Transfers, Inequality and the Poor in the Developing World, Presentation at the USAID, Washington, DC, May 15, 2014.
Maasoumi, E., 1986. The measurement and decomposition of multi-dimensional inequality. Econometrica 54 (4), 991–997.
McCall, L., 2013. The Undeserving Rich: American Beliefs About Inequality, Opportunity, and Redistribution. Cambridge University Press, Cambridge.
Meade, J.E., 1964. Efficiency, Equality and the Ownership of Property. Allen and Unwin, London.
Milanovic, B., 2013. All the Ginis You Ever Wanted. World Bank website, http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/0,contentMDK:22301380~pagePK:64214825~piPK:64214943~theSitePK:469382,00.html.
Moretti, E., 2013. Real wage inequality. Am. Econ. J. Appl. Econ. 5, 65–103.
Nolan, B., Salverda, W., Checchi, D., Marx, I., McKnight, A., Tóth, I., van de Werfhorst, H. (Eds.), 2014. Changing Inequalities and Societal Impacts in Rich Countries. Oxford University Press, Oxford.
OECD, 2014. All on Board: Making Inclusive Growth Happen. OECD, Paris.
Ostry, J., Berg, A., Tsangarides, C.G., 2014. Redistribution, Inequality, and Growth: IMF Discussion Note, SDN 14/02. International Monetary Fund, Washington, D.C.
Ovaska, T., Takashima, R., 2010. Does a rising tide lift all the boats? Explaining the national inequality of happiness. J. Econ. Issues 44 (1), 205–224.

Paine, T., 1797. Agrarian Justice, Printed by W. Adlard, Paris. Reprinted and sold by J. Adlard and J. Parsons, London.
Pigou, A.C., 1920. The Economics of Welfare. Macmillan and Co., London.
Piketty, T., 2001. Les hauts revenus en France. Grasset, Paris.
Piketty, T., 2014. Capital in the 21st Century. Harvard University Press, Cambridge, MA.
Piketty, T., Saez, E., 2003. Income inequality in the United States, 1913–1998. Q. J. Econ. 118, 1–39.
Ravallion, M., 2014. The Luxembourg Income Study. J. Econ. Inequal. (forthcoming).
Roemer, J., 1998. The Equality of Opportunities. Harvard University Press, Cambridge.
Salverda, W., Nolan, B., Checchi, D., Marx, I., McKnight, A., Tóth, I., van de Werfhorst, H. (Eds.), 2014. Changing Inequalities in Rich Countries. Oxford University Press, Oxford.
Samuelson, P.A., 1965. A theory of induced innovations along Kennedy-Weizsäcker lines. Rev. Econ. Stat. 47 (4), 444–464.
Sato, R., 1963. Fiscal policy in a neo-classical growth model: an analysis of the time required for equilibrating adjustment. Rev. Econ. Stud. 30 (1), 16–23.
Sato, K., 1966. On the adjustment time in neo-classical models. Rev. Econ. Stud. 33 (3), 263–268.
Schwabish, J.A., 2014. An economist's guide to visualizing data. J. Econ. Perspect. 28 (1), 209–234.
Sen, A., 1973. On Economic Inequality. Clarendon Press, Oxford (Expanded edition with a substantial annexe by James Foster and Amartya Sen, Oxford University Press, 1997).
Sen, A., 1976. Real national income. Rev. Econ. Stud. 43 (1), 19–39.
Sen, A., 1980. Equality of what? In: McMurrin, S. (Ed.), Tanner Lectures on Human Values, vol. 1. Cambridge University Press, Cambridge, pp. 195–220.
Sen, A., Stiglitz, J., Fitoussi, J.-P., 2010. Mismeasuring Our Lives: Why GDP Doesn't Add Up. The New Press, New York.
Social Protection Committee, 2014. Social Europe—Many Ways, One Objective. Publications Office of the European Union, Luxembourg.
Solt, F., 2009. Standardizing the World Income Inequality Database. Soc. Sci. Q. 90 (2), 231–242.
Stevenson, B., Wolfers, J., 2008. Happiness inequality in the US. J. Leg. Stud. 37, 33–79.
Stiglitz, J.E., 1969. Distribution of income and wealth among individuals. Econometrica 37 (3), 382–397.
Stiglitz, J., 2012. The Price of Inequality. W. W. Norton & Company, New York.
Summers, L., 2013. Economic possibilities for our children, pp. 1–13, The 2013 Martin Feldstein Lecture. NBER Reporter, No. 4.
Tinbergen, J., 1975. Income Distribution: Analysis and Policies. North-Holland, Amsterdam.
United Nations Economic Commission for Europe, 2011. Canberra Group Handbook on Household Income Statistics, 2nd edition. United Nations, New York and Geneva.
Veenhoven, R., 2005. Return of inequality in modern society? Test by dispersion of life-satisfaction across time and nations. J. Happiness Stud. 6, 457–487.

ACKNOWLEDGMENTS

First, we would like to thank Ken Arrow and Mike Intriligator for having persuaded us back in 2010 that the time was ripe for a new volume. How prescient their advice was!

In the course of preparing the new Handbook volume, we have incurred many debts. Without the—more than 50—authors, there would be no Handbook, and we would like to thank them sincerely for the cooperative spirit in which they have produced such excellent chapters. We much appreciate the way in which they responded to several rounds of comments from the editors and (usually) met deadlines. Part of preparation took the form of an Author's conference, entitled "Recent Advances in the Economics of Income Distribution" that was held on April 4–7, 2013 at the Paris School of Economics. We are most grateful for the support that was provided to this conference by the Agence National de la Recherche (Projet INDURA INEG 2011), Cornell University, the UK Economic and Social Research Council and DFID (Grant ES/1033114/1), Ouvrir la science économique (Opening Economics), and the Paris School of Economics.

Throughout all this, a key organizing role has been played by Véronique Guillotin of the Paris School of Economics, and we thank her most warmly.

Tony Atkinson
François Bourguignon

CHAPTER 1

The Principal Problem in Political Economy: Income Distribution in the History of Economic Thought

Agnar Sandmo
Department of Economics, Norwegian School of Economics (NHH), Bergen, Norway

Contents

Handbook of Income Distribution, Volume 2A
ISSN 1574-0056, http://dx.doi.org/10.1016/B978-0-444-59428-0.00002-3

Abstract

This chapter considers the history of theories of income distribution, from the time of Adam Smith until the 1970s. It is divided into two main parts. Part I considers the positive theory of income distribution, beginning with the classical economists' analysis of the functional distribution of income between wages, profits, and rent. It goes on to present the new theories that emerged with the marginalist revolution and which were based on maximizing behavior and market equilibrium. The main focus during the early stages of the new developments was on the markets for consumer goods and the role of marginal utility in price determination. The later neoclassical economists, including Alfred Marshall and Knut Wicksell, paid more attention to the special features that characterized the labor market and the role of marginal productivity in wage formation. In the twentieth century, the neoclassical theory was extended to include analysis of the role of imperfect competition, human capital, and risk-taking. Also included in this part of the chapter is a discussion of statistical and institutional approaches. Part II covers normative theories of income distribution and their implications for redistributive policy. It begins with a consideration of the value judgments implicit in the policy recommendations of the classical economists and continues with the attempts to establish an analytical foundation for welfare economics. The rise of Paretian welfare theory with its emphasis on the impossibility of interpersonal comparisons of utility made it difficult to draw conclusions regarding income redistribution, but the older utilitarian approach, including equal sacrifice theories, continued to live on in the modern analysis of optimal redistribution. A short Part III contains some concluding reflections on the position of income distribution theory within economics as a whole.

Keywords

Functional and personal income distribution, Distributive justice, Redistribution policy

JEL Classification Codes

B10, B20, D30, D63

1.1. INTRODUCTION

Theories of the distribution of income between individuals and classes in society have been advanced in the literature of economics from before Adam Smith to the present day. Nevertheless, although David Ricardo in the preface to his *Principles of Political Economy and Taxation* (1817; 1951, p. 5) said that the determination of the laws of distribution was “the principal problem in Political Economy,” the field has at times led a somewhat modest existence on the outskirts of mainstream academic research. One of the reasons for this may have been that the study of income distribution is so tied up with normative issues of equity and justice that many economists, keen to pursue a value-free version of their subject, have tended to shy away from it. Others, however, have found this connection to be a particularly attractive feature of the field and have risen to the challenge of clarifying the distinction between the positive and normative aspects of the analysis of income distribution, in other words, separating explanation from justification. Both aspects of the study of income distribution are reflected in this chapter, which covers the positive theory of income distribution as well as the attempts that have been made to evaluate the distribution of income from the point of view of justice and equity. The development of normative theories of income distribution is intimately tied up with the analysis of redistributive policies, which in addition also have to take account of the positive analysis of income distribution. This chapter focuses on *theories* of income distribution, while no attempt has been made to cover the large empirical literature in the area, including the statistics of income distribution. Some thoughts on the relationship between theory and empirical work in the area are presented in Section 1.4.

Taking this broad view of the field of income distribution, the literature that is relevant for this chapter becomes so large that its history cannot possibly be covered in its entirety. This is especially so because many economic theories—in areas like international trade, public finance, labor, economic growth, and so on—have implications for income distribution, although the distributive aspects are not the main concern of the researchers involved. The treatment must therefore by necessity be selective, with main emphasis on the contributions of the most important and influential economists among those who have been concerned with the theory of income distribution.

For this reason as well as for reasons of space, this chapter does not by any means attempt a complete coverage of the literature. For further references and more detailed treatment, there are fortunately a number of other sources that the reader may consult and that offer complementary perspectives on the field. They include classics like Cannan (1893, 1914) and Dalton (1920) and a number of more recent surveys like Atkinson (1975), Sahota (1978), Ranadive (1978), Asimakopulos (1987), Atkinson and Bourguignon (2000), and Goldfarb and Leonard (2005). Although not all of these have their main attention on the history of thought, they include a lot of relevant historical material. The same is true of Ravallion’s contribution to this Handbook.

All accounts of the history of thought face two fundamental questions: When to begin and when to stop. In this chapter, I have decided to start with Adam Smith, as he is arguably the first economist in whose work we begin to see the contours of modern theories. The line at the other end is drawn where the literature is still being regarded as part of the contemporary set of references. This cannot be located with a great deal of precision but has been drawn roughly at some time in the 1970s.

The chapter is divided into two main parts. Following Section 1.1, Part 2 is concerned with positive theories of income distribution, whereas Part 3 covers value judgments and redistribution; in addition, a short Part 4 contains some concluding observations. This division means that the treatment of some economists has been split in two, e.g., Pareto is discussed first in the context of the debate over Pareto's law and second in relation to his contribution to welfare economics. Although this may in some respects be unfortunate, it should be kept in mind that the main purpose here is not to give well-rounded pictures of individual economists but rather to trace the development of thought within the main areas of income distribution theory. A broader treatment of the history of economic thought, including biographical sketches of the lives of the more important economists, has been given in Sandmo (2011).

1.2. THE POSITIVE ECONOMICS OF INCOME DISTRIBUTION

It has sometimes been claimed that one of the fundamental questions that has motivated the systematic study of economics is "Why are some countries rich and some poor?" This may well be correct when we consider the motivations of some of the leading economists. But for the large majority of mankind who, at least until fairly recent times, had little opportunity to obtain firsthand knowledge of the economic conditions in foreign countries, one would have thought that a more obvious question would have been "Why are some *people* rich and some poor?" This question might naturally have come to mind as individuals went about their everyday business in a world of large inequalities of income and standard of living. On the other hand, to what extent people did reflect on this question would presumably depend on whether they thought of the inequality of income as a basic and unalterable feature of the society in which they lived or as something that followed from man-made institutions and policies that were subject to change through the political process.

It took in fact considerable time before this question moved to the forefront of economics; indeed, it may be asked whether it has ever reached the forefront. Some thoughts on this question are contained in Part 4.

1.2.1 The Classical School: Factor Prices and the Functional Distribution of Income

By the classical school of economics, we shall, in line with standard usage in the history of economic thought, refer to the economists from Adam Smith to John Stuart Mill who dominated economics during the century from the 1770s to the 1870s. The members of this school were chiefly English and Scottish, although there were also economists in Germany, France, and other countries who felt a strong affinity to Adam Smith and his successors.[1]

[1] Among the prominent followers of Smith and Ricardo in continental Europe was Jean-Baptiste Say in France. In fact, Say is the only economist outside the British Isles who is mentioned by name in O'Brien's listing of "the personnel of classical economics" (O'Brien, 2004, pp. 3–9).

Regarding the positive study of the distribution of income, the theoretical approach of the classical economists focused mainly on the *functional distribution* of income, i.e., the distribution of income between the main factors of production, and it was doubtless this distribution that Ricardo had in mind when he made his remark about "the principal problem." How these "main factors" were to be defined was of course a matter of judgment, but the classical economists saw them as being labor, capital, and land, whose incomes were wages, profits, and rent. The fact that this definition of the three main categories of income should have met with such general acceptance among economists must be seen as a reflection of the fact that this particular functional distribution represented the main class division of society in the late eighteenth and early nineteenth centuries into workers, capitalists, and landowners. Although as we shall see, there are elements in classical economic theory that go some way toward explaining the personal distribution of income, to a large extent the functional distribution was also considered an important component for the understanding of the distribution of income between persons.

The theory of the functional distribution did not, in contrast to the neoclassical theory that was developed a century later, build on a unified theoretical structure. It is therefore natural to present the theory in three parts, corresponding to the three main categories of income.

1.2.1.1 Wages

In Adam Smith's great work, *An Inquiry into the Nature and Causes of the Wealth of Nations* (1776), the first chapter presents us with his famous example of technical progress and division of labor in a pin factory. In a factory that he has seen, the complicated process of the production of a pin has been broken down into "about 18" separate operations, with the result, according to his calculations, that each of 10 men can produce 4800 times as many pins in a day as a single worker operating on his own without specialization and division of labor. One might think that this dramatic increase of productivity would lead to a corresponding increase in wages, but this is a conclusion that Smith is in fact unwilling to draw. He points out, first, that the division of labor depends on the extent of the market. Although specialization may by itself be expected to lead to higher productivity and wages, the demand side of the market limits the extent of specialization. In the highlands of Scotland, the typical farmer is often miles away from the nearest artisan and therefore has to be his own butcher, brewer, and baker, and even the artisans who are located in the small towns cannot afford to be highly specialized. Second, the mobility of labor between industries would ensure that the potential increase in the wages of the workers employed in pin production would in fact be spread thinly over the wages of workers in all industries. Third, and even more important, Smith emphasized a point that was to become a crucial component in the teaching of the whole of the classical school, viz, that any increase in the general level of wages would lead to an increase of population and therefore of the workforce, and this would tend to reverse the initial increase of wages.

This idea seems to have been part of the conventional wisdom among economic and social writers at Smith's time. In a passage that reminds one of the later work of Malthus, Smith said that "every species of animals naturally multiplies in proportion to the means of their subsistence" (Smith, 1776; 1976, p. 97). In this connection, he refers to Richard Cantillon, who in his book *Essai sur la nature du commerce en général* (1755) had argued that the standard of subsistence toward which the level of wages would gravitate must be sufficient for a working family to have four children. For experience shows, Cantillon said, that only two out of four children will be able to survive into adulthood and on average two new adults are required to ensure the reproduction of the working class.

The theory of subsistence wages received its most famous statement in the work of Thomas Robert Malthus, whose *Essay on the Theory of Population* (1798) became one of the most influential books on economics ever written.[2] Among the public at large, the book became best known for its dramatic representation of the race between population and economic progress. This was illustrated by on the one hand the natural tendency of population to grow as a geometric series, whereas food production, due to decreasing returns in agriculture, would only be able to grow as an arithmetic series. Thus, the increase of population would be held down by the shortage of food, and the income of workers would accordingly converge to the subsistence level. This was to be understood as a long run theory of wages. Malthus did not deny that wages for a limited period of time could rise above the subsistence level, but this would lead to an increase in the number of births, which over time would drive wages back to the long-run equilibrium level of subsistence.

Malthus's theory was widely accepted by the other classical economists. Gradually, however, it came to be modified regarding the essential content of the concept of subsistence. According to later thinking, a temporary increase of wages might not actually revert to the initial equilibrium level because psychological and social adaptation to a higher level of income might dampen the desire for larger families. The level of subsistence would then have to be reinterpreted as a social rather than a biological minimum amount of income, and this could well be imagined to rise over time. Technological progress, on the other hand, had no place in Malthus's view of the determination of wages.

[2] Malthus's *Essay* came out in six editions during his lifetime. The most substantial changes in its contents occurred with the publication of the second edition, which in many respects must be considered a new book. Among Malthus scholars it has therefore been common to refer to the first edition as the "First Essay" and to the second and subsequent editions as the "Second Essay."

The Malthusian theory of wages emphasized the supply side of the labor market, and little was said about labor demand. However, the reason why wages might temporarily rise above subsistence must be seen as being caused by shifts in demand, so that in an expanding economy, a series of shifts in demand might cause wages to be above subsistence even for long periods of time. The classical economists' favorite example of an expanding economy was the United States (which at the time when Smith wrote was referred to as the British colonies in North America), where the extension of the country's territory implied a continually increasing demand for labor and therefore an upward pressure on wages. The general conclusion that they drew from this example was that it was not the amount of a country's wealth that caused wages to be high; rather, it was the *growth* of the economy that was the basic cause of a high level of wages.

According to the modern way of thinking about wage determination, wages, at least in a competitive economy, are determined by the intersection of the supply and demand curve for labor. This analytical apparatus was unknown to the classical economists, but their theory can nevertheless be interpreted in these terms. The long-run equilibrium can be characterized by the intersection of a horizontal supply curve and a downward-sloping demand curve, whose position depends on the supply of other factors of production. If there is an increase in the supply of capital or land, the labor demand curve shifts to the right. In the short run labor supply is approximately inelastic, so that wages rise. But the rise in wages calls forth increased supply through an expanding population. The labor force accordingly increases until a new long-run equilibrium is reached where wages have come back to the level of subsistence, sometimes referred to as the natural price of labor. This dynamic process was described by Ricardo as follows:

> *It is when the market price of labour exceeds its natural price, that the condition of the labourer is flourishing and happy, that he has it in his power to command a greater proportion of the necessaries and enjoyments of life, and therefore to rear a healthy and numerous family. When, however, by the encouragement which high wages give to the increase of population, the number of labourers is increased, wages again fall to their natural price, and indeed from a re-action sometimes fall below it.*
>
> **Ricardo (1817; 1951, p. 94)**

1.2.1.2 Profits

Profit was regarded by the classical economists as the rate of return on capital, defined as the rate of interest plus a risk premium that varied with the nature of the capital. Actually, Ricardo gave a more general version of this definition when he stated that a capitalist would take into consideration all the advantages that one type of investment possessed over another:

> *He may therefore be willing to forego a part of his money profit, in consideration of the security, cleanliness, ease, or any other real or fancied advantage which one employment [for his funds] may possess over another.*
>
> **Ricardo (1817; 1951, p. 90)**

This is very similar to Adam Smith's theory of compensating wage differentials (to be discussed later), implying a symmetric treatment of equilibrium in the markets for labor and capital. But this broad concept of the rate of return does not in fact play much role in the work of Ricardo or any other classical economist.[3]

Although there were considerable differences among individual economists in their treatment of profits, we can still piece together a fairly unified theory from their writings. One basic question that the classical economists discussed was what it was in the working of the economic system that gave rise to a positive rate of profit. Nassau Senior (1836) provided a theory that combined the assumptions of a positive rate of time preference and the higher productivity of more roundabout methods of production. In equilibrium, capital must earn a rate of profit that compensates the investor—who is assumed to be identical to the saver—for his abstinence from current consumption. This is a formulation that foreshadows the later neoclassical theory of the rate of interest, in particular that of Böhm-Bawerk (1884–1889). In addition, the rate of profit contains a compensation for the risk undertaken by the investor. On the assumption that the investor is averse to risk, the risk premium must be positive, but because the degree of risk varies between projects and industries, the risk premium, and therefore the rate of return on capital, will show considerable variation, even assuming pure competition.

According to the classical theory, therefore, profit must be seen as the reward per unit of capital that accrues to the individual capitalist. But for a complete theory of the distribution of income from capital, one would also need a theory of the individual distribution of the ownership of capital because the income from capital accruing to the individual capitalist will be equal to the rate of return times the amount of capital owned. The determination of the ownership structure was an issue that did not receive much attention from the classical economists, and therefore their theory of the distribution of income within the capitalist class must be considered to be incomplete. On the other hand, this was an issue that did not seem to be of much concern to them. The question that formed part of Ricardo's "principal problem" was the determination of capital's share of national income, not the subdivision of this share among individual capitalists.

[3] It should be noted that there is no mention in Ricardo's *Principles* of Smith's theory of wage differentials. But this does not indicate any disagreement; Ricardo makes it clear that he limits his analysis to areas where he has something new to contribute.

1.2.1.3 Rent

Rent was the income of the landowners, defined as the rental rate per unit of land times the number of units in the possession of the individual landowner. The most influential statement of the theory of rent was contained in Ricardo's *Principles* (1817). Land varies in terms of its quality or productivity. The price of corn (Ricardo's term for agricultural produce more generally) is determined by the cost of the labor and capital required to produce a unit of corn on the land with the lowest quality, i.e., the land on the margin of cultivation. On this land rent is zero. But because the nature of the product that is grown on this land is assumed to be the same as on lands of higher quality, all corn will sell at the same price, so that a positive rent will exist on all inframarginal units of land. Rent is determined by the cost of labor and capital used on the margin of cultivation, and the position of this margin is determined by the price of corn. Therefore, Ricardo concluded, "Corn is not high because a rent is paid, but a rent is paid because corn is high" (Ricardo, 1817; 1951, p. 74). An increase in the demand for corn would imply an extension of the margin of cultivation, an increase in the labor and capital cost of production, and consequently a higher corn price. This would increase total rental income in the economy.

As in the case of profits, the theory of the functional distribution of income is of limited use when it comes to the analysis of the distribution of income within the group of landowners. An increase in the demand for corn will raise the rental rate for all landowners, but the distribution of the rental income between them will depend on the distribution of the ownership to land. On this distribution, regarding both capital and land, the classical theory is mostly silent.

What is likely to happen to the functional distribution of income in a growing economy? Ricardo's view of this issue is best explained by starting from his theory of rent. Beginning with a time when wages are above the level of subsistence, population will expand, the demand for corn will increase, and the margin of cultivation will be extended. The share of rent in national income will accordingly go up, and so will the share of labor, even after the wage rate has returned to its level of subsistence. The implication of this is that profits will fall and eventually, because of a weakening of the incentive to invest, bring the process of expansion to a halt. The economy will then have reached its stationary state, but the process toward this state may be delayed because of "improvements in machinery . . . as well as by discoveries in the science of agriculture" (Ricardo, 1817; 1951, p. 120). Thus, Ricardo saw technology as an essential determinant of the functional distribution of income, and to this would have to be added the social adaptation of the level of subsistence income if, during a process of expansion, workers became adjusted to a higher standard of living.

1.2.1.4 The Structure of Wages

In the classical theory of factor prices and the functional distribution of income, the factors of production were mostly treated as homogeneous so that the analysis could be carried out at a high level of aggregation. At the same time, it was recognized that the assumption of homogeneity was a theoretical abstraction that was particularly severe when it came to the distribution of wage income because it was obvious that wages were not in fact uniform across different professions. There could in principle be two reasons for this. On the one hand, differences in wages could be caused by competitive forces. On the other hand, they could be caused by the absence of competition, either by private restraints on the process of competition or by government regulations, the "policies of Europe," as Adam Smith used to call them.

Adam Smith's competitive theory of the wage structure is now known as the theory of compensating variations. The general idea is that wages will reflect the particular circumstances pertaining to different professions. For any particular line of work, these circumstances could be such as to imply that the wage is either above or below the average for all professions. Smith mentioned several causes of wage inequality. One of these is the "ease or hardship" of the employment. A blacksmith earns less in the course of a 12-h day than a miner does in 8 h, for the work of a blacksmith is less dirty and dangerous, and it is carried out in daylight and above the ground. Some professions are particularly honorable, and because honor is part of the reward, wages are correspondingly lower. Other professions are held in general disgrace, which has the opposite effect. The most detested of all workers is the public executioner, but relative to the hours worked, no one is better paid than he.

Smith also argued that wages will vary with how difficult and expensive it is to learn the profession, with "the constancy or inconstancy of employment," and with the amount of trust placed in the worker. His fifth and final cause of wage inequality is the probability of succeeding in one's profession. If one trains to become a shoemaker, it is virtually certain that one will be able to earn one's living by making shoes. But if one is educated as a lawyer, Smith claimed, only one in 20 will be able to do well enough to live by it. To aim at the profession of a lawyer is accordingly a lottery, and because there are so few winning tickets, these must carry very high prizes. However, the wage differences in this respect are in fact less than a rational consideration of the probabilities would imply because most people, and particularly the young, have a tendency to overestimate the probability of success. Smith suggested that this explains why so many of the young among "the common people" are ready to enlist as soldiers or go to sea.

Regarding the wage implications of education and training, Smith compared education to investment in machinery:

> *A man educated at the expence of much labour and time to any of those employments which require extraordinary dexterity and skill may be compared to one of those expensive machines. The work which he learns to perform, it must be expected, over and above the usual wages of common labour, will replace to him the whole expence of his education, with at least the ordinary profits of an equally valuable capital.*
>
> **Smith (1776; 1976, p. 118)**

This is a remarkable early statement of the main idea underlying human capital theory, which was yet to take almost 200 years to be developed more fully.

Smith's theory of the wage structure is based on the assumption of perfect competition or, in his terminology, "the system of perfect liberty." But he recognized that this was not in every respect a realistic description of actual labor markets. The guild system that regulated the entry of labor into some occupations as well as government regulations that limited the regional and industrial movement of labor could lead to wage differences that were larger than they would have been under perfect competition.

It is not entirely clear how the theory of the wage structure can be reconciled with the long-run tendency toward subsistence wages. Smith's theory of the wage structure must obviously be interpreted as one of equilibrium wage differentials. But then, if the subsistence wage is to be interpreted as the average wage, some wages must be permanently below the subsistence wage, which hardly makes sense. On the other hand, if the subsistence wage is to be understood as a long-run minimum level, it must be the case that the average wage for all workers will actually be above the subsistence level, and this conclusion is not easy to fit in with the classical theory of the long-run equilibrium theory of wages.

Smith's theory of the competitive wage structure came in for a good deal of criticism and modification by a later generation of classical economists, in particular by John Stuart Mill (1848). Mill argued that although Smith's theory might be a realistic one for the case of perfectly free competition with "employments of about the same grade" and "filled by nearly the same description of people," this case is very far from the labor markets that one actually observes:

> *The really exhausting and the really repulsive labours, instead of being better paid than others, are almost invariably paid the worst of all, because performed by those who have no choice. . . . The more revolting the occupation, the more certain it is to receive the minimum of remuneration, because it devolves on the most helpless and degraded, on those who from squalid poverty, or from want of skill and education, are rejected from all other employments.*
>
> **Mill (1848; 1965, p. 383)**

Mill concluded that Smith's hypothesis that wages tended to rise with the net disadvantages associated with different occupation was wrong, and that, on the contrary, the true relationship rather was one where "the hardships and the earnings" stood in an inverse relationship to each other. In a similar vein, John Cairnes (1874) coined the term "noncompeting groups" to describe a situation in which individuals in the labor market were prevented by lack of education and skills and the constraints imposed by their class background to compete for positions over a wide range of occupations. In other words, inequality of opportunity led to inequality of wages as well as of net advantages, i.e., wages adjusted to take account of other characteristics of the different employments.

1.2.1.5 The Laws of Distribution

We have seen that the classical economists possessed a fairly sophisticated theory of the functional distribution of income. Their theory of the personal distribution was less advanced and restricted mainly to the framework of compensating wage differentials as developed by Smith and criticized by Mill. Regarding nonlabor income, their ability to analyze the personal distribution of income was limited by the absence of a theory of the distribution of ownership. A common attitude seems to have been that the distribution of ownership to capital and land was determined by historical processes that lay outside the scope of economic science. Thus, Mill claimed that in regard to the subject of Book I of his *Principles*, which is concerned with production, the "laws and conditions of the production of wealth partake of the character of physical truths." By contrast, Book II on distribution is concerned with a subject of a quite different nature:

> *The distribution of wealth . . . depends on the laws and customs of society. The rules by which it is determined, are what the opinions and feelings of the ruling portion of the community make them, and are very different in different ages and countries. . . . But the laws of the generation of human opinions are not within our present subject. They are part of the general theory of human progress, a far larger and more difficult subject of inquiry than political economy.*
>
> **Mill (1848; 1965, p. 200)**

It is clear from the context that Mill meant this statement to apply to all aspects of the distribution of income and wealth. However, he was also careful to emphasize that although the causal factors behind the distribution of income had to be studied in a broad context, including noneconomic considerations, the *consequences* of different distributional arrangements "must be discovered, like any other physical or mental truths, by observation and reasoning."

1.2.1.6 The Marxian Perspective

The basic structure of Karl Marx's positive economic theory is consistent with the teaching of the classical economists, especially Smith and Ricardo. As in their work, his main interest in the theory of income distribution lay in the functional distribution of income and less in the distribution of income between persons. He adopted the theory of subsistence wages but added an additional component, which was absent in the work of Smith and Ricardo, viz, the existence of unemployment. According to Marx, even the subsistence level of wages would not be low enough to secure full employment in the capitalist system, and the result of this was the development of what he named "the industrial reserve army" of the unemployed who live in extreme poverty and misery. He also argued that the existence of this reserve army is in fact in the interest of the capitalists. The reason is that there are significant fluctuations in economic activity that also imply large fluctuations in the demand for labor. The reserve army serves as a depository of labor on which the capitalists can draw without having to bid up wages, which they would have been led to do in a situation of full employment. Inequality and poverty therefore serve the interests of the ruling class, i.e., the capitalists.

Marx emphasized strongly that a central feature of the capitalist system was its ability to accumulate capital and generate economic growth. So what happens to the reserve army of the unemployed with the accumulation of capital? There are two effects that work in opposite directions. On the one hand, a more capital intensive technology increases the productivity of workers and tends to push wages up. On the other hand, the new technology also increases industrial concentration, and this effect lowers labor demand and pushes wages down. In the context of an increasing population, the net result of these effects may well be that employment increases, but the industrial reserve army will also increase, both in absolute and relative terms:

> *The greater the social wealth, the functioning capital, the extent and energy of its growth, and, therefore, also the absolute mass of the proletariat and the productiveness of its labour, the greater is the industrial reserve army. The same causes which develop the expansive power of capital, develop also the labour-power at its disposal. The relative mass of the reserve army increases therefore with the potential energy of wealth.*
>
> **Marx (1867-1894; 1995, pp. 360–361)**

According to Marx, therefore, and in sharp contrast to the view commonly held by the classical economists, unemployment was a permanent feature of the capitalist economic system and was central for a proper understanding of the distribution of income and wealth.

Apart from the emphasis on unemployment, a central concept of Marx's analysis of the distribution of income is *exploitation*. At the bottom of this concept is the view that labor is the fundamental factor of production in the sense that all nonlabor inputs can be derived from past labor: "As values, all commodities are only definite masses of congealed labor time" (Marx, 1867-1894; 1995, p. 16). The worker's productivity is a reflection of his labor-power. But he is only paid the subsistence wage, which is less than the value of what he produces. The difference between the two is the worker's unpaid work for the benefit of the capitalist. This is the profit or surplus value that defines the capitalist's exploitation of the worker.

Regarding the distribution of income from capital, a central element in Marx's theory is the tendency—or the law, as he calls it—of the rate of profit to fall as capital accumulates. The effect of this would be to diminish the importance of capital income. On the other hand, Marx also believed that this would go together with increasing concentration in industry and a strengthening of the monopoly element in capital income, and this would serve to counteract the first effect. Monopoly also explains his emphasis on absolute rent in addition to the Ricardian differential rent. Absolute rent arises because the absence of competition in landed property prevents rent from being brought down to zero on land at the margin of cultivation.

Marx did not limit himself to the presentation of his arguments in terms of abstract reasoning but also provided vivid examples of the living conditions in contemporary industrial society, above all in England where he lived during the last three decades of his life and where he wrote *Capital*. In this he was also able to draw on the insights and knowledge of his friend and collaborator Friedrich Engels. Engels's study of the conditions of the English working class (Engels, 1845) provided important material for Marx's own work, but is also a significant contribution in its own right. Engels, who worked as a manager in an industrial firm in Manchester that was partly owned by his father, was appalled by the living conditions of the workers that he saw in the industrial towns in England. In his book, he attempted to give a detailed description of their incomes, housing, and health, arguing that at least at this stage of the Industrial Revolution, workers were worse off than they had been before. He based his work both on his own observations and on various contemporary reports, and the book is notable for its extensive use of statistical data to describe social and economic conditions among the working-class poor.

1.2.2 Neoclassical Economics: The Marginalist Approach to the Distribution of Income

The marginalist revolution and the birth of neoclassical economics marked a new style of economic theorizing in which, in contrast to the classical writers, the new generation of economists attempted to anchor their analysis in the behavior of individual economic agents, using the theory of optimization and the mathematical tools of the differential calculus. But it also marked a new view of the workings of the market economy. Particular stress has traditionally been laid on the greater attention to demand as a determinant of prices, but the differences were also substantial when it came to the study of income distribution. To a large extent, the development of a new approach to income distribution was driven by the internal logic of theoretical innovation, but there can be little doubt that it was also motivated by the social and economic development that became increasingly visible toward the end of the nineteenth century. As an example we may take Léon Walras, who criticized Malthus for the lack of logic in his theory of population, in particular for his neglect of the role of technological progress. He also pointed out the failure of Malthusian theory to explain the actual increase in living standards for all classes in society. Thus, after having been impressed by the progress demonstrated at the World Exhibition in Paris in 1867, he wrote an article where he emphasized the benefits that advances in technology had brought to the working class and confronted them with the "ridiculous theory" of Malthus, predicting the workers' eternal poverty and misery.

1.2.2.1 The Marginalist Revolution and Its Forerunners

Although the marginalist revolution is usually identified with the early 1870s, there were important forerunners of neoclassical economics who in some respects were actually more advanced in their analytical approach than their successors. Foremost among the early champions were Johann Heinrich von Thünen and Herrmann Heinrich Gossen in Germany and Antoine Augustine Cournot and Jules Dupuit in France. In the present context, it is von Thünen and Gossen that have a special claim to our attention.

Von Thünen's main work *Der Isolierte Staat* (*The Isolated State*, 1826, 1850) is remarkable in this connection particularly for his early formulation of marginal productivity theory, which he applied both to capital and labor use. Thus, for a producer who attempts to maximize profits, he derived the conditions that the value of the marginal productivities of labor and capital must be equal to the wage rate and interest rate, respectively, and he used this approach to study geographical variation of the choice of capital intensity in a spatial economy. Von Thünen considered the result of equality between marginal value productivities and factor prices also to be a theory of income distribution, but as such it is obviously incomplete in that it takes no account of the supply side of factor markets, thus leaving the formation of factor prices unexplained (except for the special case where factor supplies are given). Nevertheless, this was an important building block for the theory of factor prices that was to be developed later.[4]

Hermann Heinrich Gossen's long-neglected book on economic theory (Gossen, 1854) is famous mainly for its early formulation of the theory of the utility-maximizing consumer and its derivation of "Gossen's law" that at the optimum the ratio between marginal utility and price must be the same for all consumer goods. In the central version of his theory, income is taken as given so that it does not include any theory of factor supply, but he did in fact present an extension of his model in which he claims that the supply of labor can be derived from the condition that the marginal utility of consumption is equal to the disutility of work. Together, von Thünen and Gossen provided important elements for the theory of factor price formation and income distribution, but it was yet to take a long time before their approach had been developed into a logically consistent theory of income distribution.

What historians of economic thought commonly refer to as the marginalist revolution is associated with three authors and three books: William Stanley Jevons's *Theory of Political Economy* (1871), Carl Menger's *Grundsätze der Volkswirtschaftslehre* (1871), and Léon Walras's *Eléments d'économie politique pure* (1874–1877). The central concern of the three main protagonists of the marginalist revolution in the 1870s was to establish the theory of subjective value as the main causal factor for the understanding of price formation. This led them to focus first of all on the determination of prices for consumer goods, but they also extended the theory to apply to the formation of factor prices. The equality of marginal value productivities and factor prices as following from profit maximization is

[4] Von Thünen has become particularly famous for the formula for "the natural wage," which is equal to the square root of the product of the existence minimum and worker productivity. There is general agreement that in the history of thought this should be treated as a curiosity rather than a substantive contribution (although von Thünen thought sufficiently highly of it to have it inscribed on his gravestone).

particularly explicit in Walras (1874–1877; 1954, Lesson 36). Walras also emphasized that a theory of the average rate of wages—which he considered to be the main focus of the classical economists—is not very useful; the analysis of wages must be based on a disaggregated view of the labor market with occupation-specific wage rates. However, neither Walras nor the other two went very far in the analysis of income distribution. Although they considered the application of the marginalist method to the analysis of wages and interest rates, they did not proceed to a study of how the theory could be used to explain inequality in society. For this we have to wait for the work of a later generation of marginalist or neoclassical economists, and in the coming decades, a number of writers made important contributions. Here, we shall focus on the work of Alfred Marshall and Knut Wicksell, who both in different ways left their mark on the development of economics during the next century.

1.2.2.2 Alfred Marshall

The contrast between the work of Léon Walras and Alfred Marshall has frequently been characterized as that between general and partial equilibrium theory. That is clearly true regarding their style of theoretical analysis. But in addition, it is striking how much their great treatises differ with regard to the reliance on institutional and empirical material. Thus, when Marshall approached the issue of what determines the demand for labor, he did it by way of a numerical example in which a sheep farmer decides how many shepherds to hire at a given rate of wages, hiring more workers as long as an additional shepherd's marginal value product exceeds the wage rate. He emphasized that the theory that "the wages of every class of labor tend to be equal to the net product due to the additional labor of the marginal laborer of that class" does not in itself constitute a complete theory of wages because a number of other aspects both of factor and product markets need to be taken into account.[5] On the other hand, "the doctrine throws into clear light the action of one of the causes that govern wages" (Marshall, 1890; 1920, p. 518).

As Walras before him, Marshall also argued that phrases such as "the general rate of wages" were apt to be misleading, for

> *... in fact, there is no such thing in modern civilization as a general rate of wages. Each of a hundred or more groups of workers has its own wage problem, its own special set of causes, natural and artificial, controlling the supply-price, and limiting the number of its members; each has its own demand-price governed by the need that other agents of production have of its services.*
>
> **Marshall (1890; 1920, p. 533)**

There is an interesting contrast here to the work of Adam Smith and John Stuart Mill in that the wages of labor are analyzed from the start within the framework of multiple

[5] For a discussion of the relationship between the concepts of net and marginal product as used by Marshall, see Whitaker (1988). For the case of perfect competition and full substitutability of the factors of production the two concepts coincide.

(although interrelated) labor markets, whereas the classical economists discussed the general rate of wages, later adding on a somewhat ad hoc discussion of wage differentials. The supply and demand framework instead provided a general approach to the study of wage formation, which could be used to analyze both the general level of wages (assuming, contrary to Marshall, that there is such a thing) and the wage differentials between occupations. However, Marshall also discussed the theory of compensating wage differentials, blending elements from the partially conflicting views of Smith and Mill.

Although Marshall must clearly be considered to be one of the founding fathers of the marginal productivity theory of wages,[6] his theoretical perspective was much wider than this terminology may indicate. Among his significant theoretical innovations in the study of wages and the distribution of labor income should be counted his early formulation of the theory of human capital. He noted that

> *[t]he professional classes especially, while generally eager to save some capital* for *their children, are even more on the alert for opportunities of investing it* in *them.*
>
> ***Marshall (1890; 1920, p. 533)***

Although investment in children by means of education and training will increase their productivity and thereby their opportunity to earn good wages, there are some serious imperfections in the market for human capital. One of these is the weakness of employers' incentives to invest in human capital. This capital becomes the property of the worker, so that the employer's opportunities of reaping the gains of any investment that made in the worker is severely limited, hence arises the crucial role of the parents, which is limited by "their power of forecasting the future, and by their willingness to sacrifice themselves for the sake of their children" (Marshall, 1890; 1920, p. 561). But although the parents play an important role in overcoming the adverse incentives of employers, this role has also other and more unfortunate consequences. Because the opportunities and insights of the professional classes are not shared by the members of the "lower ranks of society," their investment in their children is inadequate, and this evil is cumulative:

> *The worse fed are the children of one generation, the less will they earn when they grow up, and the less will be their power of providing adequately for the material wants of their children; and so on to following generations.*
>
> ***Marshall (1890; 1920, p. 562)***

Another point that Marshall repeatedly stressed is the dependence of productivity on wages. High wages lead workers to be better fed and better educated and so increase their productivity. Marshall suggested that this mechanism may be an important part of the

[6] This term has become a standard one among historians of economic thought, although Marshall himself would no doubt have objected to it as being an incomplete description of his own theory of wage formation.

explanation of the historical increase in wages, contrary to the predictions of at least the simple version of the Malthusian theory.

Both his emphasis on a disaggregated view of the labor market and his early insistence on the importance of human capital and efficiency wages make Marshall a very important contributor to the theory of income distribution, at least as regards the distribution of labor income. About the distribution of income from capital he has less to say. He applied marginal productivity theory to the study of the rate of interest, but because he did not offer any theory of the distribution of the ownership of capital (and land), the distribution of income from capital becomes an unsolved issue. The contrast to labor income is an interesting one: Because the discussion of the marginal productivity of labor is usually framed in the context of man-years of labor (as in the shepherd's example), and because the measurement of the distribution of labor earnings uses annual income as its basis, the distribution of wages becomes identical to the distribution of earnings. Thus, the marginal productivity theory becomes a much more important element in the theory of the distribution of labor income than in the study of the distribution of income from capital.[7]

1.2.2.3 Knut Wicksell

The Swedish economist Knut Wicksell is an important figure in the history of the marginalist revolution and the rise of the neoclassical school of economic theory. Whereas the earlier marginalists—apart from von Thünen—had focused most of their attention on the analysis of consumption, Wicksell's main interest was in production and investment decisions. It is worth noting that his initial interest in economics was kindled by his concern for social problems and the issues raised by unchecked population growth. In Volume 1 of his *Lectures on Political Economy* (1901–1906), he argued that virtually every problem in economics had to be studied in the context of a changing population; however, the population issue in fact plays relatively little role in his more formal academic writing.

Wicksell is especially well known for the first clear and precise formulation of the production function as a central tool in the analysis of production and investment decisions (including the original introduction in economics of what became known as the Cobb–Douglas function). He made explicit the idea of factor substitution, and the assumption of continuous substitution between factors of production was adopted by later economists as a defining characteristic of neoclassical economics. In a more rigorous fashion than his contemporaries, he showed that profit maximization involved the equality between marginal value products and factor prices. Like Marshall, he stressed the incompleteness of marginal productivity theory as a theory of income distribution because it did not take the supply side into account. He did not really manage to

[7] The shortcomings of marginal productivity theory in explaining the distribution of income from capital and land were strongly emphasized both by Cannan (1893) and Dalton (1920).

integrate supply and demand in a formal analysis of income distribution, but in his discussion of practical issues, he showed a clear understanding of the nature of their interaction. Although he emphasized the role that technological progress had played in increasing the marginal productivity of labor, he held the view—in sharp contrast to Walras—that it was doubtful whether real wages had shown any increase during the preceding 200 years, whereas rent in his opinion had "successively doubled and redoubled." The explanation for this he found in the growth of population during the same period:

> *Such an increase [in population] must, other things being equal, continually reduce the marginal productivity of labour and force down wages; or—what comes to the same thing, though the connection is easily overlooked on a superficial view—*prevent *the otherwise inevitable* rise *in wages due to technical progress.*
>
> ***(Wicksell (1901-1906; 1934, p. 143)***

As a purely theoretical proposition, this statement shows a very clear understanding of the respective roles played by supply and demand in the determination of wages. On the other hand, its empirical connection with actual economic developments during Wicksell's lifetime is highly questionable and can only be interpreted as being strongly colored by his neo-Malthusian convictions.[8]

A further important theoretical issue in the neoclassical analysis of production and distribution concerns the problem of product exhaustion: Would the payments to the factors of production according to marginal productivity theory exhaust the value of output? Earlier, Philip Wicksteed (1894) had shown with reference to Euler's theorem of homogeneous functions that this would happen if firms' production functions were linearly homogeneous. The problem with this application of the theorem was that it implied constant marginal and average cost, so that the scale of production for each firm was indeterminate. Wicksell pointed out that the problem would be solved by the assumption that production functions went through phases of increasing, constant, and decreasing returns to scale. This corresponds to the case of an average cost function, which first decreases and then increases. At the minimum point of the U-shaped cost curve there are constant returns to scale, and this is in fact the point to which the long-run equilibrium of the industry will converge, given the assumption of free entry. Factor prices correspond to marginal value productivities, and the payments to the factors of production exhaust the value of the product with pure profits being zero. But even in the case where product prices are given, as when they can be taken to be determined in world markets, this theory of distribution is incomplete in the absence of a theory of factor supply.

[8] For a more general discussion of the relationship between theory and statistical evidence in the work of income distribution theorists, see Part 4.

1.2.2.4 General Equilibrium Theory

The work of the neoclassical economists—from that of the early pioneers to the first and second generation of the marginalists in the closing decades of the nineteenth century—became consolidated in the later version of the theory of general equilibrium that was developed around the middle of the next century. The main achievements of this development have often been associated with the introduction of new methods of mathematical methods in economics and with the analysis of existence and stability of equilibrium, but in a broader perspective one must also include the deeper understanding of the general interdependence in the economy that it led to. A particularly important aspect of this interdependence was the relationship between the prices of consumer goods, factor prices and the distribution of income and wealth. But the connection between resource allocation and the distribution of income was not given much attention in modern general equilibrium theory; in the influential presentation of the theory by Debreu (1959), the term *distribution* does not even appear in the index. In one respect, however, the modernized version of the Walrasian system provided a more satisfactory treatment of distribution. Dalton (1920) had criticized the marginal productivity theory of distribution for not giving a satisfactory account of the distribution of income from capital and land. The theory treated only the determination of the rate of interest and the rent from land, but the distribution of capital and rental incomes had to be concerned with the interest rate *times* the ownership of capital and with the rental rate *times* the holdings of land.[9] This shortcoming of the theory is resolved in the modern theory by the introduction of the notion of endowments. Consumers are assumed to be endowed with initial resources (in principle both consumer goods and factors of production) as well as shares of the profits of the different firms in the economy, so that prices do indeed determine the distribution of income or wealth. On the other hand, part of Dalton's criticism remains valid because endowments and profit shares are taken to be exogenous, and no account is provided of their origin.

One reason why the new mathematical formulation of general equilibrium theory paid little explicit attention to the problem of income or wealth distribution was that in its ambition to achieve a high degree of generality, it rid itself of the distinction between consumer goods and factors of production. Formally, consumer goods were defined as commodities that entered the budget constraints as positive numbers, whereas factors of production were commodities represented by negative numbers. Moreover, the focus of the theory was on the competitive case, so that there was no scope for treating the formation of factor prices, e.g., wages, as being any different from the formation of

[9] Cannan (1893) had directed a similar criticism against the classical economists, calling the functional distribution of income with which they were chiefly concerned a "pseudo-distribution" because it was only concerned with wages per head, profits per cent, and rent per acre.

prices for consumer goods. Labor was just like any other commodity and wages no different from all other prices.

In applications of the general equilibrium framework the situation was different. In international trade theory, the effect of international trade on the domestic distribution of income had long been a central focus of the theory, and in the 1940s and 1950s, the analysis of the connection between the prices of factors and goods moved to the forefront of the theoretical development in the field; the classic contributions were Stolper and Samuelson (1941) and Samuelson (1953). The focus of this literature was on the functional distribution of income, in particular on the shares of labor and capital, whereas the analysis of the personal income distribution was mostly by implication, as in the study of sectoral shifts following changes in world market prices.

Another field in which one might expect the general equilibrium framework to be important for the study of income distribution is public economics. But this has hardly been the case. One explanation for this is that in contrast to international trade theory, public economics has always had a strong concern with the effect of taxes on factor supply, whereas in international economics one has often been content with assuming factor supplies to be given. The extension of the framework of analysis to incorporate variable factor supply leads to significant complications, and this may be the main reason why the best-known use of the general equilibrium approach in public economics is Harberger's (1962) analysis of the incidence of the corporation income tax. Harberger's model turned out to be a fruitful one for analyzing a number of problems in tax incidence analysis. On the other hand, the reason why it was easy to use was precisely because, in analogy with international trade theory, it ignored the study of the effects of taxation on the supply of capital and labor, issues that have otherwise been treated as central in the theory of public economics.

1.2.2.5 Imperfect Competition

The early neoclassical economists and the later general equilibrium theorists focused their analysis of the market economy on the case of perfect competition. In the case of the labor market, the assumption was that both workers and employers took the equilibrium market wage as given, whereas the forces of competition made any out-of-equilibrium wage rate adjust until the supply of labor was equal to demand. It was within this framework that theorists discussed the dual role of wages—and more generally of factor prices—in allocating factors of production among alternative uses and determining the distribution of factor incomes.

That the case of perfect competition was not a realistic one particularly in the labor market was already acknowledged by Adam Smith in his discussion of the determinants of wages (Smith, 1776, Book I, Chapter VIII). He emphasized that wages are influenced both by private and public restraints on competition. The guild system limits the access to certain occupations and thereby pushes up the level of wages relative to that of other lines of employment, and the government tolerates these regulations. Another point that he makes is that in bargaining over an employment contract, the natural advantages are with the employers. There are fewer employers than workers, so that it is easier for the employers to collude to keep wages low than it is for workers to combine to push wages up. Smith wrote long before the time of strong trade unions, and he remarked that although there are many laws that forbid workers to organize themselves for the purpose of obtaining higher wages, there are none that prevent employers in colluding for the opposite purpose. He also pointed out that if a conflict occurs, the employers can hold out much longer than the workers. A factory owner will often be able to live well without workers for a year or two, whereas a worker will find it difficult to survive for a week or a month if not employed. The implication is evidently that in many labor markets wages will be lower than they would have been in a situation of perfect competition with bargaining power being symmetrically distributed.

It took a long time before Smith's insights were taken into account in the neoclassical theory of the market economy. Pigou's *Economics of Welfare* (1920) discussed the functioning of the labor market with careful attention to the role of various institutions that interfere with competition in one way or the other. Because the relationship of the parties in the labor market is one of imperfect competition, there is an unavoidable indeterminateness in regard to the level of wages. In Appendix III to his book (Pigou, 1920; 1952, pp. 813–814), he provided a diagram that shows the deviation of the equilibrium wage from the competitive level,[10] but he did not attempt to identify exactly what determines the imperfectly competitive wage level.

The year 1933 saw the publication of the two books that moved the concepts of monopolistic and imperfect competition into the core of economic theory. *The Theory of Monopolistic Competition* by Edward Chamberlin had its focus on the markets for consumer goods, whereas Joan Robinson's *Economics of Imperfect Competition* also contained an analysis of imperfectly competitive labor markets with obvious implications for the distribution of income (which, however, she did not discuss except in passing). Pigou's indeterminateness was removed by the assumption of completely asymmetric bargaining power by the two parties to the labor contract: Employers were assumed to be monopsonists, and workers took wages as given. This led to an equilibrium in which wages were in general below the level of the marginal value products, with the gap between them reflecting the elasticity of supply. The larger the value of the elasticity of supply, the smaller would be the gap between the two, and the less would be the degree of exploitation. The implications of imperfect competition in the labor market were also considered by Hicks (1932), whose book among a number of other issues also contained an extensive discussion of the role of trade unions. In regard to the theory of income

[10] He also used the deviation between the perfect and imperfect competition level of wages to measure what he called unfairness and exploitation.

distribution, however, Hicks's main interest was in the functional rather than the personal distribution of income. Thus, one of his most influential contributions in the book was the analysis of the effects of various types of technical progress on labor's share of national income.

The general indeterminateness of the outcome of wage bargaining, which was stressed by Pigou, also played a central role in the theory developed by the Danish economist Frederik Zeuthen in his book *Problems of Monopoly and Economic Warfare* (1930).[11] His theory is set in the framework of a bilateral monopoly model in which a firm bargains with a trade union and where neither party has any outside option; the employer has no alternative use of his capital, and workers have no alternative employment opportunities. While recognizing the basic indeterminacy of the equilibrium solution, Zeuthen explored the factors that would determine the features of the bargaining process and the likely outcome. Both parties realize that failure to reach agreement will result in a conflict—a strike or a lockout—that will be costly to both of them. Zeuthen saw the bargaining process as a series of proposals and counterproposals, where proposals of high wages would make employers willing to risk a conflict, and this would put downward pressure on wages. Proposals of low wages, on the other hand, would make the union more willing to risk a conflict and thereby tend to push wages upward. At some intermediate wage level, both parties will consider the risk of pushing for a better alternative to be equally large, and this will be the equilibrium wage. Zeuthen's theory was an important contribution to better understanding of the role of bargaining and labor conflicts and a significant extension of the neoclassical theory of labor markets and income distribution.[12]

1.2.2.6 Human Capital Theory

An unsatisfactory aspect of the marginal productivity theory of distribution—quite apart from its neglect of the supply side of factor markets—was that it offered little explanation of why some factors of production were more productive than others. One might argue that this was simply a question of technology and the way that factors were combined in the production process, but particularly in the case of labor, it is hard to escape the belief that some individuals are in some sense inherently more productive than others. However, some of the differences in productivity might be due to education and training. This point was already made by Adam Smith, and we have also seen that Alfred Marshall suggested a possible explanation for this in the investment that parents made in their children, both with the time that they themselves devoted to them and with the resources that they

[11] Actually, the theory had been presented two years earlier in his doctoral dissertation, published in Danish (Zeuthen, 1928), which is a broad theoretical and empirical study of the income distribution in Denmark.

[12] As pointed out by Harsanyi (1955a), it was also, together with the analysis by Hicks (1932), a forerunner of the game theoretic approach to bargaining associated with John Nash (1950).

spent in giving the children a good education. This would result in higher wages for the children who benefited but possibly also in increased inequalities of wage income.

Another writer who pursued the idea of investment in human beings was the German statistician Ernst Engel. In his 1883 book on the cost value (Kostenwerth) of human beings, he calculated the cost of training a boy to practice his father's profession in the lower, middle, and upper classes of society (corresponding to lower, middle, and higher education).[13] However, he did not have a theoretical framework that allowed him to explore the analogy between investment in human and physical capital, and he did not discuss the implications of his approach for the distribution of income, implicitly ruling out the possibility of mobility between income classes.

In the twentieth century, the ideas of Smith and Marshall were taken up by the economists of what came to be called the human capital school. Although important contributions were made by Theodore Schultz (1961), the theoretical foundations were laid by Gary Becker (1962, 1964). In particular, Becker's 1964 book marked the beginning of an extremely influential line of research, which also took up important issues regarding the distribution of income. As set out in Becker and Chiswick (1966), the amount of investment in human capital at the individual level is determined by the intersection of the supply and demand curve (or the marginal benefit and the marginal cost curve). Both supply and demand curves must be expected to vary among individuals. Different supply curves may reflect the income and wealth of parents and access to capital markets, whereas the position of the demand curve may represent individual characteristics like inherent ability and attitudes to risk. In Becker and Tomes (1979), the framework is extended to an intergenerational setting where children's endowments are partly determined by the investments made in them by their parents. This is clearly related to the ideas of Marshall regarding the long-term effects of investment in children.

As with all theoretical innovations, the growth of the human capital field can to some extent be explained by developments internal to the discipline of economics. However, it is also natural to point out explanations that reflect changes in the economy. Studies of economic growth had led to increased attention to changes in the efficiency of labor as a determinant of growth. Perhaps, more to the point in the present connection are the consequences of an increasing level of education in the labor force, which made the distinction between income from capital and labor seem a less central element in a realistic theory of income distribution. A society in which an increasing number of workers had become human capitalists required a new perspective on the distribution of income.

[13] Engel also considered the costs of education for girls, but in their case, he did not include a calculation of the cost of higher education.

1.2.2.7 Risk Taking and Income Distribution

The difference of riskiness of income between occupations figured as one element in Adam Smith's theory of compensated wage differentials. In the choice between a safe and a risky occupation (shoemaker and lawyer in Smith's example), the expected wage in the risky occupation would have to be higher than in the safe one to compensate individuals for their additional risk bearing. To the extent that individuals assessed the probabilities correctly, these *ex ante* expectations would be translated into *ex post* income inequality: The incomes of lawyers would have a higher average but greater variance than the wages of shoemakers.

The possibility of formal modeling of choice in risk-taking situations was greatly stimulated by the axiomatic foundation of expected utility theory developed by von Neumann and Morgenstern (1947). Although it took some time for the theory to find applications in the analysis of real economic problems, its use in the theory of income distribution was one of the earliest. The classic article in the field is by Milton Friedman (1953), who used his earlier work with Leonard Savage (Friedman and Savage, 1948) to explain income distribution as the result of rational choice under uncertainty. A distinctive feature of the Friedman–Savage theory is the assumption that they make about attitudes to risk. Although the assumption of risk aversion is a natural one for explaining real-world features like portfolio diversification and insurance, it does not explain the simultaneous existence of gambling. To resolve this difficulty, Friedman and Savage assumed that the utility function of income had both concave and convex segments, i.e., ranges of both decreasing and increasing marginal utility. In Friedman's income distribution theory, individuals at the beginning of their lives choose between alternative income streams; at the level of abstraction of Friedman's analysis, these streams could be generated from labor as well as capital income. Although individuals have equal opportunities *ex ante*, the income lotteries in which they engage imply that some will find themselves *ex post* with high incomes, and some will end up in low-income groups. The special shape of the utility function gives rise to a distribution of income that, Friedman argued, is consistent with observed patterns, in particular as documented in his own empirical work with Kuznets (Friedman and Kuznets, 1945). He also argued that individuals will be motivated as participants in a democratic society to introduce redistributive mechanisms that insure them against the consequences of the most adverse outcomes. According to this theory, therefore, both income inequality and redistributive policies emerge as results of individuals' free choice in a situation of equality of opportunity and will reflect their attitude to risk, in particular the relative importance of risk averters and risk lovers. The less risk-averse individuals are, the greater will be the inequality of income in society.

A further development of this framework is due to Kanbur (1979), whose analysis builds on a much more specific structure than that used in Friedman's article. In Kanbur's framework, risk-averse individuals choose between the safe occupation of a worker and the risky occupation of an entrepreneur. In equilibrium, the two occupations must be equally attractive, i.e., have the same expected utility, and this implies that the expected income of the entrepreneur must be higher than that of the worker. Kanbur explored the comparative statics of the model and showed that when account is taken of general equilibrium effects on the distribution of individuals between occupations, there is no longer any simple connection between risk aversion and inequality. In a companion paper, Kanbur (1981) studied the role of taxation in the determination of the equilibrium distribution of the population between the two occupations.

On this point, Kanbur's study is related to the older analysis of taxation and risk taking that goes back to the classic article by Domar and Musgrave (1944). Their analysis of a model of portfolio choice showed that under certain assumptions, particularly that of full loss offset, income taxation induces individuals to take more risk than they otherwise would have done. Their choice of more risky portfolios obviously has the implication that their wealth *ex post* will have a larger variance than it would have had in the absence of income taxation.[14] With full loss offset, income taxation functions in part as insurance against variations in capital income, and this insurance acts as an encouragement to risk taking. *Ex post*, therefore, one would expect higher taxation to generate more inequality in the distribution of income from capital.

1.2.3 Nonmarginalist Approaches

The marginalist revolution of the 1870s left its mark on the style of economic theorizing for a long time; indeed, it remains a dominating influence on contemporary economics. As we have seen, it also played a central role in the theory of income distribution. But at the same time, other contributions were made that do not easily fit into the marginalist framework. A common feature of the alternative approaches is that they pursued an inductive rather than a deductive line of investigation. Some of these will be discussed later.

1.2.3.1 Statistical Approaches: The Pareto Distribution

Although the marginalist theory held out the promise of a theoretically more firmly based theory of the personal distribution of income, the late nineteenth century also saw the introduction of a more inductive theory of income distribution, founded not on *a priori* theorizing but on inference from statistical data. The pioneering contribution was due to Pareto, whose work caused a good deal of discussion and controversy during several decades after its initial publication.

Vilfredo Pareto was Walras's successor in the chair of economics at the University of Lausanne. Like Walras, he was a firm believer in the mathematical method, and he saw it

[14] The Domar–Musgrave article did not use the expected utility hypothesis. For a reformulation and sharpening of their theory along expected utility lines, see Mossin (1968).

as his main task to extend and refine the general equilibrium approach that Walras had developed, including the theory of factor price formation. When it comes to income distribution, however, Pareto's fame rests not on his refinements of Walrasian theory but on his formulation of what has become known as Pareto's law.[15] Many economists only know Pareto from footnotes in textbook treatments of utility theory and welfare economics and may be forgiven for thinking of him as a pure theorist. But Pareto was an immensely productive researcher who wrote on a wide variety of topics, both theoretical and empirical, and not only in economics. He is a significant figure in the history of sociology and wrote also on statistical theory, economic history, and political science. His studies of income distribution, set out in a number of articles and in his book *Cours d'économie politique* (Pareto, 1896–97) drew on his knowledge both of economics and mathematical statistics and, in the matter of interpretation, also on his insights in sociology.

What posterity has come to know as Pareto's law was not derived from a theoretical model; instead, it was based on a detailed study of incomes statistics for a number of countries and time periods. Pareto's analysis of these data led him to the hypothesis that all statistical income distributions have a common shape that one can characterize as follows. Suppose that we draw up a list of all incomes in society from the lowest to the highest. Starting from the median income, we know that 50% of the income earners have an income above the median. We then move up to a level of income that is 1% higher than the median and ask what percentage of the population has an income above this level. Obviously, the percentage is <50, but how much less? Pareto found that the answer was 1.5%; in other words, as the level of income goes up by 1%, the number of individuals with an income above this level falls by 1.5%. In general, mathematical terms Pareto wrote his law as:

$$\log N = \log A - \alpha \log y.$$

Here, N is the number of individuals who have an income of at least y, and A is a parameter that reflects the size of the population. α is Pareto's constant that he estimated to be approximately equal to 1.5. The relationship has the interesting property that the average income of those whose incomes are greater than y will be equal to $\alpha/(\alpha-1)$ times y. Thus, once again assuming that $\alpha=1.5$, the average income of those with incomes above 10,000 francs should be equal to 30,000 francs. In the economies that Pareto studied, it turned out that the fit of the function was remarkably good, although less so at the lower tail of the income distribution. Later work has tended to establish that the fit is particularly good for the upper ranges of the distribution, i.e., for the right end of the income distribution curve.

[15] His other claim to fame is of course his role in the development of welfare economics, which will be considered later.

Pareto's law came in for a good deal of controversy. Thus, a long discussion involving several participants arose regarding Pareto's claim that the parameter α could be used as an index of inequality. That this claim should turn out to be controversial will come as no surprise to the modern economist, who from the work of Atkinson (1970), Sen (1973), and others has been made aware that any particular index of inequality is implicitly based on some ethical judgment about the nature of inequality. The question of the conditions required for social welfare to be written as a function of mean income and inequality as measured by Pareto's α (increasing in the former, decreasing in the latter) was settled by Chipman (1974). Having this issue clarified is of obvious interest. However, there were other aspects of the controversy that are arguably of greater general importance.

One question that naturally arises concerns the empirical validity of the law. Did Pareto actually claim the law to be one of universal validity? Here, his statements do not provide an unambiguous answer. On the one hand, he noted in a comment on his empirical findings that

> *[t]hese results are very remarkable. It is absolutely impossible to assume that they are due solely to chance. There must certainly be a* cause *which produces a tendency for incomes to be distributed along a certain curve. The form of this curve seems to depend only slightly on different economic conditions of the countries considered, since the effects are about the same for countries in which economic conditions are as diverse as those of England, Ireland, Germany, Italian cities and even Peru.*

However, he went on to issue a word of caution:

> *True, since we are dealing only with empirical laws, we cannot be too prudent. In any case, the consequences we shall draw from this law will at least always be valid for peoples for whom we have seen that they are confirmed.*
>
> ***Pareto (1896-97), vol. II; quoted from Chipman (1976, p. 151)***

Despite this and other cautionary statements Pareto was frequently interpreted as claiming universal validity for his law. Such a claim naturally proved provocative to many who believed that governments should see it as one of their objectives to bring about a more egalitarian distribution of income. On the one hand, Pareto seemed to claim that the distribution of factor incomes was given; on the other hand, he also went out of his way to point out that, given the skewness embedded in the Pareto distribution of incomes, progressive taxation could only be counted on to provide a rather insignificant redistribution of income in favor of the poor. This was seen by many as proof of Pareto's alleged reactionary attitudes, although this view is not supported by statements such as

> *. . . even with taxes at an equal percentage of incomes, the rich contribute far less to public expenditures than the poor, whereas they benefit much more from them. For whom, if not for the vain rich, are funds expended on armaments and the like?*
>
> ***Pareto (1895); quoted from Chipman (1976, p. 115)***

However, it was the early presentation of Pareto, rather than his later and more cautious statements, that caught the attention of other economists, and a considerable amount of work was devoted to examining and criticizing his law of income distribution. Thus, in his *Economics of Welfare* (1920), Pigou devoted a whole chapter (Part IV, Chapter II) to a critical examination of Pareto's law. In the preceding short chapter, Pigou had sketched the principles underlying the equity–efficiency trade-off (to use a more modern expression), arguing from a utilitarian perspective that any cause that increases the "national dividend" without lowering the absolute share of the poor, or increases the absolute share of the poor without reducing the national dividend, must increase welfare. By contrast, the welfare effect of any measure that increases one of these quantities but diminishes the other is ambiguous:

> *Plainly, when this kind of disharmony exists, the aggregate effect upon economic welfare, brought about by any cause responsible for it, can only be determined by balancing in detail the injury (or benefit) to the dividend as a whole against the benefit (or injury) to the real earnings of the lower classes.*
>
> **Pigou (1920; 1932, p. 645)**

Pigou then went on to point out that, according to one "interesting thesis," there was no need to be concerned about these cases of disharmony: Pareto's alleged law of income distribution implied that because the relative shares of the different income groups were at least approximately constant, the only way to ensure an increase in the absolute share of the poor was to increase the national dividend. Pigou was clearly skeptical of the conclusion and also expressed strong doubts with respect to several aspects of Pareto's work. He criticized the empirical basis for Pareto's generalization, but a more important point that he raised concerns the basis for assuming a given distribution relating to all sources of income. Pareto's distribution is skewed to the right, and Pigou argued that in the case of labor income one would rather like to assume that the distribution of "capacities" follows the normal distribution.[16] He also pointed out, however, that capacity is a multidimensional concept, and that although manual and mental capacity might both be normally distributed, their joint distribution would not be, and this fact might go some way toward explaining the form of the Pareto distribution. On the other hand, the reference to capacity, whether manual or mental, does not explain the distribution of income from capital or property, which is largely determined by inheritance, the importance of which depends in a crucial manner on the nature of legal and political institutions. The view

[16] If capacity is taken to mean marginal productivity it is, of course, not sufficient to argue that the normal distribution of capacity is reflected in a corresponding normal distribution of wages. According to marginal productivity theory, wages correspond (under competitive conditions) to the value of the marginal products, so that the distribution of wages also depends on the distribution of product prices and accordingly on the distribution of workers between industries.

that the distribution of income, and in particular the share of the poor, cannot be affected by measures of economic policy therefore becomes untenable.

Toward the end of the chapter Pigou quotes Pareto as remarking about his own distribution that

> *[Some] persons would deduce from it a general law as to the only way in which the inequality of incomes can be diminished. But such a conclusion far transcends anything that can be derived from the premises. Empirical laws, like those with which we are here concerned, have little or no value outside the limits for which they were found experimentally to be true.*
>
> **Pigou (1920; 1932, p. 655)**

So it appears that Pigou's criticism of Pareto to some extent missed its target. That it still was felt to be necessary to devote a chapter to it in 1920 must be explained by the popular attention that Pareto's original formulation had attracted. The idea that the distribution of income was determined by a sort of immutable law appeared to have far-reaching consequences for the feasibility—or rather infeasibility—of redistributive policies.

Pigou was not the only economist to be critical of Pareto's law of income distribution. Edgeworth (1896) at an early stage of the debate argued that Pareto's contribution bore strong similarities to previous work by the English statistician Karl Pearson. Pareto reacted strongly to what he saw as an accusation of plagiarism and gave a heated reply in which he remarked that "it must have displeased Mr. Edgeworth to see me poach on territory which is apparently reserved for Professor Pearson, just as political economy is reserved for Professor Marshall" (Pareto, 1896). Further exchanges did little to soften the tone of the debate, and as late as 1926, 3 years after Pareto's death, Edgeworth wrote about Pareto's reaction that it "is of interest as throwing light not only on the character of the curve, but also on that of its discoverer" (Edgeworth, 1926; 2003, p. 492).

Pareto's formulation of his law as well as the later controversies to which it gave rise constitute an interesting episode in the history of economic thought, and the Pareto distribution continues to play a role in the empirical study of income distribution. Although it has received a good deal of criticism, it has also been hailed as a milestone in the empirical study of income distribution.[17]

1.2.3.2 Other Statistical Approaches

The tradition established by Pareto's work to look for regularities or empirical laws in the distribution of income was continued by a number of later writers. A characteristic feature of this literature is that the authors do not attempt to found their hypotheses on the neoclassical theory of factor market equilibrium but start instead from

[17] For a survey of the statistical literature that, although critical, takes an overall positive view of Pareto's contribution, see Bresciani-Turroni (1939). A balanced survey of the controversy surrounding Pareto's law is the article by Chipman (1976).

some observed empirical regularity, just as Pareto did. Just a few examples of this approach will be given here.

Roy (1950, 1951) claimed that observed earnings distributions could be reasonably approximated by the lognormal distribution and argued, echoing Pareto, that "[t]here must be some rational explanation of the fact that all these earnings' distributions have such similar shapes" (Roy, 1950, p. 490). He attempted to discover this explanation by studying a number of industrial cases in which workers performed a standard and identical task and where individual output was easy to measure. These included tasks like packing boxes of chocolate, stitching shoes, and pressing gramophone records. Altogether, for the 12 different cases studied, it turned out that the lognormal distribution performed slightly better than the normal. To the extent that people are paid according to output, this result could go some of the way toward explaining the earnings distribution in terms of the distribution of individual skills. In Roy (1951), he studies the theoretical case of a "primitive" society in which people can choose to work in two or more occupations and where their skills differ between occupations. He then discussed how different skill correlations give rise to different statistical earnings distributions (always assuming that earnings are proportional to output), emphasizing the central role played by the lognormal distribution. Champernowne (1953) considered a dynamic model in which it is assumed that every income earner has a probability of a rise or fall in income between one period and the next, which is proportionate to his income in the first period. He showed that over time this will result in convergence toward the Pareto distribution. In a comment on this article, Lydall (1959) argued that this stochastic process was implausible for labor incomes and showed that the Pareto distribution could be generated on the alternative assumption that in an industrial firm each supervisor controls the same number of persons and is paid according to the total income of those below him. A similar assumption about the pyramidal structure of organizations was employed by Herbert Simon (1957) in his analysis of the compensation of executives.

A different and more macroeconomic approach was taken by Kuznets (1955), whose goal was to explain the long-term trends in the inequality of income in the economy as a whole. Although on the basis of data for the United States, England, and Germany, he found that income inequality had decreased after the end of the First World War, he suggested that this period had been preceded by one of increasing inequality. In his view, the period of widening income gaps began with the Industrial Revolution in the late eighteenth century; for England he suggested that it ended around the middle of the nineteenth century and for the others a few decades later.[18] His explanation for this development was based on the shifts from the agricultural or traditional sector of

[18] Setting the date of the change from the first to the second phase at roughly 1850 for England, Kuznets suggested that Marx's view of the inevitable rise of inequality of income under capitalism may have been an overgeneralization from observations of the last stages of the first phase.

the economy to the nonagricultural or modern sector, where income from capital plays a larger role for the distribution of income. Initially, inequality is larger in the modern sector than in the traditional one, and this generates an increased inequality of income for society as a whole as the modern sector expands. Over time, however, as the modern sector becomes more mature a variety of forces combine to reduce inequality there, particularly through an increased share of the lower-income groups and a lowering of the income from capital. Consequently, overall inequality diminishes. In his own words:

> *One might thus assume a long swing in the inequality characterizing the secular income structure; widening in the early phases of economic growth when the transition from the pre-industrial to the industrial civilization was most rapid; becoming stabilized for a while; and then narrowing in the later phases.*
>
> **Kuznets (1955, p. 18)**

This hypothesis is what has become known as the Kuznets curve in the form of a bell-shaped curve describing the relationship between per capita income and the degree of inequality. It should be emphasized, however, that Kuznets was careful to point out the inadequacy of the empirical evidence for the hypothesis, particularly in regard to the earlier phase of economic growth.

The various statistical approaches to the study of income distribution are attempts to rationalize the observed distribution of income by using some stylized facts or assumptions about the generation of income to explain observed patterns of the distribution of income. To call these approaches, nontheoretical might be somewhat misleading; however, it is clearly the case that they are not founded on theories of optimizing behavior and market equilibrium.

1.2.3.3 Institutional Theories of Income Distribution

There have always been economists who were skeptical of the central role played by formal models in economic theory. In the area of income distribution, we have seen that even a prominent theorist like John Stuart Mill argued that "the laws of distribution" must be understood in a political and social context, and because this context was determined by institutions, the understanding of the distribution of income and wealth would have to take proper account of institutions in addition to the mechanism of demand and supply. Karl Marx emphasized that the distribution of income in the society of his time reflected the particular phase of social development that he called capitalism. Along similar lines, the German historical school, led by Wilhelm Roscher and Gustav Schmoller, downplayed the role of theory in favor of an approach based on a detailed study of historical data. If successfully carried out, this line of research would presumably be less able than, e.g., the marginal productivity theory to offer explanations with a claim to universal validity; on the other hand, it might hold out a promise of generating more insights with relevance for the particular society being studied.

It was especially in the United States that institutional approaches to the study of the economic system received a position that made many regard it as an important alternative to the theoretical approach of the neoclassical school of economists. Thorstein Veblen is widely regarded as the founder of American institutional economics, but his approach—more satirical than analytical—in books like *The Theory of the Leisure Class* (1899) and *The Theory of Business Enterprise* (1904) was too idiosyncratic to attract many direct followers.[19] Neither he nor the other most prominent members of the institutional school, John R. Commons and Wesley C. Mitchell, paid particular attention to the distribution of income except for a general emphasis on the importance of power relations and evolutionary processes. The chief importance of the institutional school may have been as critics of the neoclassical theory in its focus on rational behavior and competitive equilibria. But the lack of general propositions in the work of the institutional school contributed to its gradual decline as an influence on modern economics.

An interesting question that arises in the study of the effects of institutions on the economy is: What constitutes an institution? Here, Veblen adopted a broad definition that encompassed "settled habits of thought common to the generality of men." A modern version of this idea came with Gary Becker's work on the economics of discrimination (Becker, 1957), in which racial discrimination in the labor market is assumed to arise from a common preference for not working alongside people with a different skin color. In pursuing the implications of this idea, Becker may be said to have followed the guidelines for economic research recommended by the institutional economists; however, the tools that he used in this work were entirely neoclassical.

Regarding the inequality of wage income, important contributions have been made by specialists in labor economics and industrial relations. It is natural to group these with the institutional economists because like them they emphasize the crucial role of institutions for the understanding of the distribution of income, specifically the distribution of wage income. In the United States, the work of Dunlop (1944, 1958) described wages as determined by the interaction between company owners, management, and workers as represented by trade unions.[20] The book by Phelps Brown (1977) collects a number of his studies of wage inequality in different countries and under different economic systems.

[19] The closest that one may come to such a follower is perhaps John Kenneth Galbraith, whose satirical style and skepticism toward mainstream economics are in many ways reminiscent of Veblen. His book *The Affluent Society* (1958) contains several discussions of issues of income distribution with criticism of mainstream views but does not offer any alternative explanations of observed patterns of inequality.

[20] It should be noted that Dunlop's work is not institutional in the sense of showing aversion to theoretical modeling. As an example, in his 1944 book, he discussed the formal mathematical modeling of trade union behavior in a situation of unemployment, analyzing the relationship between the union's wage claim and the rate of unemployment compensation and thereby the distribution of income between the employed and the unemployed. This analysis foreshadows the numerous contributions to the theory of trade union behavior in the 1970s and 1980s.

His work is notable for the attempt to explain inequality of pay by drawing both on economic and sociological approaches, paying attention to such factors as social class and status, discrimination, intergenerational mobility, and mental ability.

1.2.3.4 The Role of Property Ownership and Inheritance

The role of inheritance as a determinant of income distribution has received relatively little attention in the theoretical literature. In the world of the early neoclassical economists and the later general equilibrium theorists, the subject did not fit easily into their models. The time dimension—essential to get a grip on inheritance—could indeed be added through the introduction of time-dating consumer goods as well as factors of production, but this failed to provide a convincing picture of the nature of inheritance. In the world of general equilibrium theory, as described, e.g., in the book by Arrow and Hahn (1971), property ownership was represented by "endowments," initial holding of goods and factors of production that were taken as exogenous. But models of this type are unable to explain the passing on of property from parents to children and the persistence of inequality between generations. The nature of these intergenerational transfers is determined by the rules of inheritance, which will therefore have an important influence on the distribution of income and wealth. But as Dalton remarked almost a century ago,

> *Many thinkers of high reputation still talk, or remain silent, about the law of inheritance, as though it had fallen immutable from heaven into the Garden of Eden.*
>
> **Dalton (1920, p. 285)**

Meade (1964) considered the development of the personal distribution of wealth on the background of what he saw as the likely development of the functional distribution of income. In his view, the dominating technological trend was toward "automation," which would imply a significant reduction in the demand for labor and falling wages. This would lead to a shift in the functional distribution of income away from labor and in favor of income from property. Because, as he pointed out, income from property is much more unequally distributed than income from labor, this shift would imply a greater overall inequality in the population. This trend toward increased inequality in the distribution of income might in Meade's view be reinforced by demographic factors, such as higher rates of growth for large than for small fortunes (due to better opportunities for diversification), the genetic inheritance of earning power, and the tendency toward assortative mating (the rich marrying the rich). As later pointed out by Stiglitz (1969), it could also be influenced by the rules governing inheritance, either by law or custom. If all wealth goes to the firstborn (primogeniture), this leads to a more unequal distribution of wealth than the alternative of dividing wealth equally among one's children.

Inheritance is of obvious importance not only for material wealth but also for human capital. We have seen that this point had already been emphasized by Marshall (1890), and some decades later Cannan argued that the individual qualities required both to earn a good income from labor and to manage one's property wisely were passed on from one generation to the next, so that this tended to stabilize the degree of inequality over time. However, this tendency was not without exceptions:

> *The able members of the poorest class are constantly rising to the top, and the particularly incompetent members of the richest class are constantly falling to the bottom; but all the same, among the bulk of mankind there is a continuous hereditary transmission of inequality of income, the importance of which it is foolish to ignore.*
>
> **Cannan (1914; 1928, p. 217)**

The role of inheritance in determining the degree of inequality in the ownership of property is obviously an important one and requires attention to the broader subject of what Mill called "the laws and customs of society." Perhaps, his warning, that this was a much larger and more difficult subject than economics, played some role in the development that led economists largely to neglect this important aspect of the distribution of income and wealth.

1.3. VALUE JUDGMENTS AND REDISTRIBUTION

The interest in the question "Why are some people rich and some poor?" has always been motivated by something more than pure intellectual curiosity. A notable feature of the observed distribution of income has always been that it is unequal, and a natural second question is therefore "Can inequality be justified?" A possible response to this question is that it is one that should be answered by moral philosophers and not by economists, whose science does not provide them with the tools needed to answer it. There are indeed some economists who have taken this position, but there are also a large number who have not, and this includes many of the most prominent characters in the history of the subject. The reasons for this are not difficult to see. On the one hand, there is the fact that many economists—from Adam Smith to Amartya Sen—have had a foot in the camp of the moral philosophers, so that crossing the borders between the two fields has come naturally to them. On the other hand, there is the existence of the borderland between the two fields, which is the study of the effects of redistribution policy. To understand the design and consequences of redistribution policy, one must know something both about economics and moral philosophy, and the attempts to combine them constitute the normative part of the study of income distribution.

1.3.1 The Normative Economics of the Classical School

The natural starting point for economic theories of distributive justice is the distribution of income that is generated by the market economy. Although the main concern of the classical economists was with the positive analysis of income distribution, they were also concerned with ethical issues and with the evaluation of redistribution policy.

1.3.1.1 Adam Smith

A point of reference for the classical view of this issue is Adam Smith's theory of the invisible hand. In the most famous single passage in the *Wealth of Nations*, he claims that each individual, by pursuing his self-interest also promotes the interest of society:

> *He intends only his own gain, and he is in this, as in many other cases, led by an invisible hand to promote an end which was no part of his intention. Nor is it always the worse for the society that it was no part of it. By pursuing his own interest he frequently promotes that of the society more effectually than when he really intends to promote it.*
>
> **Smith (1776; 1976, p. 456)**

The most common interpretation of this passage is that private incentives operating in the context of a market economy promote an efficient use of resources in the sense of maximizing "the annual revenue of society," although this interpretation is not undisputed.[21] Does it also promote a just distribution of income? There is no systematic discussion of this in the *Wealth of Nations*, although most readers of the book will find it reasonably clear that this was not his view. It is remarkable, therefore, to find in Smith's other main work, *The Theory of Moral Sentiments* (1759), a paragraph in which he makes the claim that the rich, without intending to do so, promote the interests of the poor. His statement of this claim is also of interest because it contains the second of his three uses of the metaphor of the invisible hand.[22] The rich, he says

> *... in spite of their natural selfishness and rapacity, though they mean only their own conveniency, though the sole end which they propose from the labours of all the thousands whom they employ, be the gratification of their own vain and insatiable desires, they divide with the poor the produce of all their improvements. They are led by an invisible hand to make nearly the same distribution of the necessaries of life, which would have been made, had the earth been divided into equal portions among all its inhabitants, and thus without intending it, without knowing it, advance the interest of the society.*
>
> **Smith (1759; 1976, pp. 184–185)**

The proposition that the distribution of necessaries is almost the same as if the economic system had been designed with a view to an equal distribution is certainly a striking one, although one should note that there is no claim that the income that finances the consumption over and above that level is distributed in a similar fashion. The self-interest of the rich is claimed to guarantee a certain minimum income to the poor, but not to the extent of leading to equality of living standards. Almost regardless of one's interpretation of the substantive content of this proposition, it is difficult to see that Smith provides any convincing support for it, and it is hardly surprising that this version of the invisible hand has had little influence on subsequent thinking about income distribution.

[21] For a discussion of alternative interpretations of the meaning of Smith's statement of the invisible hand, see Chapter 2 of Sandmo (2011).

[22] The third use occurs in his essay on the history of astronomy.

Going back to *The Wealth of Nations*, although it does not contain any systematic discussion of the normative aspects of the distribution of income, there are many passages in the book that demonstrate Adam Smith's concern with inequality and poverty as well his sympathy for the poor. One example is his positive attitude toward trade unions, which leads him to suggest that it is an inconsistency of economic policy to allow employers to collude while forbidding workmen to form trade unions (Smith, 1776; 1976, pp. 83–85). Another example that, although in itself of minor importance, is suggestive of his attitude, is his discussion of the system of the tolls that should be charged for different types of public transport. The principle that was most commonly used at Smith's time was that of charging according to the weight of the carriage. He argued against this principle and in favor of the alternative of charging higher rates for luxury carriages and lower rates for carriages of necessity. Such a reform, he argued, would have the effect that "the indolence and vanity of the rich is made to contribute in a very easy manner to the relief of the poor, by rendering cheaper the transportation of heavy goods to all the different parts of the country" (Smith, 1776; 1976, p. 725).

A clearer statement of Smith's more general perspective on the distribution of income between rich and poor comes in a passage that follows a discussion of the effects of lower prices of necessities:

> *Is this improvement in the circumstances of the lower ranks of the people to be regarded as an advantage or as an inconveniency to the society? The answer seems at first sight abundantly plain. Servants, labourers and workmen of different kinds, make up the far greater part of every great political society. But what improves the circumstances of the greater part can never be regarded as an inconveniency to the whole. No society can surely be flourishing and happy, of which the far greater part of the members are poor and miserable. It is but equity, besides, that they who feed, cloath and lodge the whole body of the people, should have such a share of the produce of their own labour as to be themselves tolerably well fed, cloathed and lodged.*
>
> ***Smith (1776; 1976, p. 96)***

It is clear from the context that Smith meant this statement to apply even to the case where the improvements in the standard of living of the lower ranks were achieved at some cost to the higher ranks of society.

What consequences did Smith draw for redistributive policy? Here, we must keep in mind that the instruments available for redistributive policy were limited in number in Smith's time, so that his policy recommendations were mostly incidental, as in the preceding passage concerning charges for public transport. His discussion of taxation in Book V of the *Wealth of Nations* is not very explicit when it comes to the redistributive effects of the tax system as a whole; he is content to discuss the main categories of taxes one by one with apparently little regard for the overall impact of the tax system. However, this discussion is introduced by the presentation of four normative "maxims" of taxation, and in the first of these we find the following principle:

> *The subjects of every state ought to contribute towards the support of the government, as nearly as possible, in proportion to their respective abilities; that is, in proportion to the revenue which they respectively enjoy under the protection of the state. In the observation or neglect of this maxim consists what is called the equality or inequality of taxation.*
>
> **Smith (1776; 1976, p. 825)**

The principle may not be entirely clear to the modern reader and could be interpreted in two different ways. The first part of the passage indicates that the principle is one of ability to pay, whereas the second part might suggest that we should read it as a recommendation of the benefit principle, according to which taxes should be seen as payment for services rendered by the state. However, the most reasonable interpretation of the term "revenue" is "income"; a central service that the state provides is security of private income, so that income is both a measure of ability to pay and benefits received. Thus, the tax system as a whole should be as nearly as possible proportionate to income. It is important to note that this is not a recommendation for the form of an income tax—about which Smith has little to say—but for the more general design of the tax system as a whole.

1.3.1.2 Malthus and Ricardo on the Poor Laws

Although redistributive taxation played little role at the time of the early classical economists, the form that support for the poor should take was a major issue of public policy.[23] There was widespread concern over the established system of poor relief, which provided assistance both to those too sick or too old to work and to those who were able to work but found it difficult or impossible to earn a living. Malthus applied his theory of population to this issue and argued that support for the poor would not in the long run improve their position in society. Because the provision of a minimum standard of living would encourage the poor to have more children, in the long run they would not be better off on an individual basis; there would simply be a larger number of poor people in society. In addition, the resulting increase of population would drive up the price of food and cause more workers to rely on poor relief:

> *They [the poor laws] may be said, therefore, to create the poor which they maintain; and as the provisions of the country must, in consequence of the increased population, be distributed to every man in smaller proportions, it is evident that the labour of those who are not supported by parish assistance will purchase a smaller quantity of provisions than before, and consequently more of them must be driven to apply for assistance.*
>
> **Malthus (1803; 1992, p. 100)**

Malthus therefore recommended the abolition of the poor laws to increase the incentives of the able-bodied poor to provide for themselves through their own work. In this he

[23] The history of thought regarding public policy toward the poor is discussed both more broadly and in more depth in Martin Ravallion's chapter in the present *Handbook*.

received strong support from other prominent economists, in particular from his friend David Ricardo. According to Ricardo, "the comforts and well-being of the poor" cannot be secured without some effort of their own, especially to regulate the increase in their numbers. But, he argued,

> *The operation of the system of poor laws has been directly contrary to this. They have rendered restraint superfluous, and have invited imprudence, by offering it a portion of the wages of prudence and industry.*
>
> *The nature of the evil points out the remedy. By gradually contracting the sphere of the poor laws; by impressing on the poor the value of independence, by teaching them that they must look not to systematic and casual charity, but to their own exertions for support, that prudence and forethought are neither unnecessary nor unprofitable virtues, we shall by degrees approach a sounder and more healthful state.*
>
> **Ricardo (1817; 1951, p. 107)**

In a stark form the critique of the poor laws introduced a theme that was destined to become a major issue in the economic analysis of poverty and redistribution: The possible conflict between the objectives of justice (poor relief) and efficiency (labor supply). Later classical economists, in particular Nassau William Senior, who was chairman of the 1832 Royal Commission on the poor laws, strongly recommended a reform of the system that ensured that poor relief would never be organized in such a way as to make it more attractive than to earn one's living by regular work.

1.3.1.3 Mill

John Stuart Mill is known as one of the most prominent spokesmen for the philosophy of utilitarianism, which he expounded in particular in his book *Utilitarianism* (1863). One might expect then that in his *Principles* he would use the utilitarian approach to evaluate income inequality, but this perspective is in fact absent from his analysis.[24] Like in the case of Adam Smith, we search in vain for a unified theoretical principle that can be used to evaluate income distribution from a normative point of view. On the other hand, there are numerous opportunities to gain insight into his views on distribution from his discussion of more specific issues.

On such issue is that of inheritance. Although Mill supports each individual's rights to the fruits of his own labor and property, he draws a line when it comes to income from inherited property. In a passage that may have been more controversial to his readers than he indicated (Mill, 1848; 1965, p. 218), he wrote that "although the right of bequest, or gift after death, forms part of the idea of private property, the right of inheritance, as distinguished from bequest, does not." He therefore supported restrictions regarding inheritance in the form of limits on how much an individual may be allowed to receive. His

[24] The last edition of the *Principles* that appeared during Mill's lifetime was the 7th, which came out in 1871. Thus, he clearly had the opportunity to use material from *Utilitarianism* for this purpose.

arguments for such restrictions run partly in the form of incentives: Although restrictions on how much a parent is allowed to leave to his children may weaken the parent's desire to accumulate wealth, this is outweighed by the adverse incentives to work and save that arise for children who receive large amounts of wealth that they have done nothing to deserve. But he also defended the proposed restrictions by its distributional consequences. If children's inheritance were to be limited to some maximum amount,

> *. . . the benefit would be great. Wealth which could no longer be employed in "over"-enriching a few, would either be devoted to objects of public usefulness, or if bestowed on individuals, would be distributed among a larger number.*
>
> ***Mill (1848; 1965, p. 226)***

According to Mill, therefore, there is a social benefit associated with a more even distribution of wealth.[25]

Another issue is that of the most desirable form of taxation. In his chapter "On the General Principles of Taxation" (Mill, 1848; 1965, Book V, Chapter II), Mill cited with approval Adam Smith's four maxims on taxation. After having quoted them *in verbatim*, he commented that although their meaning is mostly clear, the maxim that is concerned with equality in taxation (and which was cited earlier) requires further examination because it is concerned with a concept that is often imperfectly understood. He then stated that the fundamental principle of equality in taxation is equality of sacrifice, which means ". . . apportioning the contribution of each person toward the expenses of government, so that he shall feel neither more nor less inconvenience from his share of the payment than every other person experiences from his" (p. 807). He then went on to discuss the consequences of this general principle for the design of the income tax. Although expressing some sympathy for the idea of a graduated income tax, he concluded in favor of a linear tax in which, e.g., the first 50 pounds of income is tax exempt, whereas the excess income is taxed at a constant rate. He also recommended that saving be exempt from taxation, the main argument being that taxing the parts of income that are devoted to consumption and saving at the same rate involves a "double taxation of saving" and therefore a disincentive to saving and investment.

Mill's tax policy recommendations emerge as a compromise between the abstract idea of equal sacrifice and more *ad hoc* considerations, but it is difficult to see to what extent his conclusions can be derived from the philosophical principles of utilitarianism. In his book *Utilitarianism* (Mill, 1863; 1969, pp. 254–255), there is a brief discussion of alternative concepts of justice in taxation, but the text is rather inconclusive: Mill described alternative points of view that give support to a head tax, a proportional tax, or progressive taxation. He then stated that "[f]rom these confusions there is no other mode of

[25] A century later, Mill's recommendations were echoed by Meade (1964), who proposed progressive taxes both on wealth and inheritance for the purpose of achieving a more equal distribution of the ownership of property.

extrication than the utilitarian." However, he did not conclude as to the form of taxation that would follow from the application of utilitarian principles, and as we have seen this connection is not clear in his discussion in the *Principles* either.

It may seem surprising that John Stuart Mill, an intellectual known for his radical sympathies, should not have come out more strongly in favor of redistributive taxation. The main explanation is probably that he saw taxation as being of secondary importance in this regard in comparison to structural reforms aiming to expand the range of choice open to all layers of society. Such reforms would include better education for the lower classes, ending the restrictions on entry into various occupations as well the discrimination of women in the labor market. The latter issue was one that he considered to be of special importance. He wrote the influential book *On the Subjection of Women* (Mill, 1869), and in the *Principles* he wrote,

> *Let women who prefer that occupation [as a wife and mother]; adopt it, but that there should be no option, no other* carrière *possible for the great majority of women, except in the humbler departments of life, is a flagrant social injustice.*
>
> ***Mill (1848; 1965, p. 765)***

It is notable that it was to take more than a century for the gender issue once again to make its appearance in the normative economics of inequality and income distribution.

1.3.2 The Neoclassical Economists: Efficiency and Justice

With the emergence of marginalism and the neoclassical school of economic theory there began a more systematic exploration of the optimality properties of the market allocation of resources and in particular the relationship between on the one hand the efficiency of the market economy and on the other hand the distributive justice of its allocation of resources. In the long-run perspective of the history of ideas, the neoclassical interest in these issues may be seen as a desire to clarify Adam Smith's proposition that the invisible hand of the market led to a result that was in conformity with "the public interest."

1.3.2.1 Walras

The three main protagonists of the marginalist revolution paid little attention to the role of the competitive market system in the determination of income distribution and even less to the ethical aspects of it. Among the three, however, Léon Walras is notable for raising an issue that goes back to Adam Smith's theory of the invisible hand and the ability of the market mechanism to function in a way that is consistent with the public interest. Toward the end of his detailed analysis of exchange in a two-commodity world he wrote that

> *[the] exchange of two commodities for each other in a perfectly competitive market is an operation by which all holders of either one, or of both, of the two commodities can obtain the greatest*

> *possible satisfaction of their wants consistent with the condition that the two commodities are bought and sold at one and the same rate of exchange throughout the market.*
>
> **Walras (1874-1877; 1954, p. 143)**

The context makes it clear that Walras meant the conclusion to apply beyond the simple case of two commodities and pure exchange, so it must be understood as a more general characterization of a competitive economy.

The characterization can be read as a modernized version of Smith's statement about the invisible hand; however, it can be interpreted in two different ways. Several economists have taken the view that the expression "the greatest possible satisfaction of their wants" refers to the collective society of all individuals; according to this interpretation, Walras said that the competitive equilibrium generates the greatest possible satisfaction of wants for society as a whole. In this perspective, Walras came out as a rather naïve apologetic for the free market system. The other interpretation is obviously that each individual can obtain the greatest possible satisfaction of wants *for himself*. There can in fact be no doubt that the second interpretation is the correct representation of Walras's position. On the one hand, he insisted that his analytical description of the competitive market has no broader normative significance:

> *Though our description of free competition emphasizes the problem of utility, it leaves the question of justice entirely to one side.*
>
> **Walras (1874-1877; 1954, p. 257)**

On the other hand, he emphasized the noncomparability of utility, so that he must have rejected the notion that there exists such a thing as wants satisfaction for society as a whole.

On the latter point, however, we have evidence that for Walras, at least in this case, old habits of thought died hard. In a letter to the German economist Wilhelm Launhard in 1885 Walras defended himself against the charge that he had maintained that competition necessarily led to maximum satisfaction for society as a whole. Suppose, he argued, that commodities can be sold at a low price to the poor and a high price to the rich. The rich would then have to give up some consumption of "superfluous" goods, whereas the poor would be better able to afford necessities. "Consequently, there would be a large increase in utility" (Jaffé, 1965, vol. II, p. 50). Here, utility evidently refers to aggregate or social utility; hence, there is an assumption, contrary to the statement in the *Eléments*, that individual utilities can be compared and aggregated.

In addition to this lapse from theoretical consistency, the modern economist might also question Walras's use of the example of price discrimination for consumer goods to illustrate redistribution policy. Clearly, an example that would both be more striking and more realistic would be redistribution of *income* from the rich to the poor. The consequences in terms of the consumption of luxuries and necessities would be the same, and the connection with policies that were within the realm of the feasible would be much stronger.

In modern terminology, the conclusion to which Walras came close, although he did not manage to state it with great clarity, was that the market equilibrium was efficient although it did not necessarily result in a just distribution of resources and income. Although imperfectly formulated, this insight was a step forward in the understanding of the connection between the market mechanism as a system for efficient resource allocation and as a determinant of the distribution of income and welfare between individuals in society. The insight was to be further studied and clarified by the next generation of marginalist thinkers of whom the most important were Alfred Marshall and Walras' successor in Lausanne, Vilfredo Pareto.

1.3.2.2 Marshall

What were Marshall's views regarding the normative aspects of income distribution? In welfare economics, Marshall is chiefly remembered for his invention of the partial equilibrium concept of the social surplus (the sum of producers' and consumers' surplus), which can be measured as the area between the demand and marginal cost curves. Because this area achieves its maximum at the point of intersection between the two curves, i.e., at the competitive equilibrium, Marshall was able to conclude that

> *a position of (stable) equilibrium of demand and supply is a position also of* maximum satisfaction*.*
>
> **Marshall (1890; 1920, p. 470)**

This is a conclusion very similar to that of Walras, although Marshall was more careful in qualifying it to avoid misunderstandings. It is obvious that he meant the conclusion to apply beyond the simple case of an individual commodity to the general equilibrium of demand and supply, including the markets for the factors of production. And although the term *maximum satisfaction* was meant to apply to society as a whole, Marshall emphasized that it is an aggregate measure that is built on the assumption that

> *all differences in wealth between the different parties concerned may be neglected, and that the satisfaction which is rated at a shilling by any one of them, may be taken as equal to one that is rated at a shilling by any other.*
>
> **Marshall (1890; 1920, p. 471)**

He then argued that if, e.g., it were the case that the producers as a class were much poorer than the consumers, "aggregate satisfaction" might be increased by a restriction of supply that would, assuming demand to be inelastic, increase the income of the producers. The terminology here is apt to be confusing because it seems strange to argue that aggregate satisfaction can be increased by moving away from a position of maximum satisfaction. But quite apart from the terminology, the underlying argument is clearly based on the utilitarian assumption of decreasing marginal utility:

> *It is in fact only a special case of the broad proposition that the aggregate satisfaction can* primâ facie *be increased by the distribution, whether voluntarily or compulsorily, of some of the property of the rich among the poor.*
>
> **Marshall (1890; 1920, pp. 471–472)**

In his concluding chapter on "Progress in relation to the standards of life" he became at the same time more explicit and more cautious regarding the desirability of less inequality:

> *The drift of economic science during many generations has been with increasing force towards the belief that there is no real necessity, and therefore no moral justification for extreme poverty side by side with great wealth. The inequalities of wealth though less than they are often represented to be, are a serious flaw in our economic organization. Any diminution of them which can be attained by means that would not sap the springs of free initiative and strength of character, and would not therefore materially check the growth of the national dividend, would seem to be a clear social gain.*
>
> **Marshall (1890; 1920, pp. 713–714)**

This is a forceful expression of the view that excessive inequality is a social evil, and one notes also Marshall's claim that this moral judgment can claim the support of economic science. On the other hand, the desirability of a move toward increased equality must take account of the possibility that it might weaken productivity and economic incentives, a point of view that would become a cornerstone in the analysis of welfare state policies that was to occupy the work of many economists in the coming generations.

What would be the means that could be used to achieve reduced inequality? On this topic Marshall's *Principles* has less to contribute. There is the emphasis on education as a means of improving one's position in society but little attention to the possibility of compulsory redistribution that he alludes to. Foremost among the instruments of such redistribution is taxation, but there is hardly any systematic discussion of the principles of taxation in Marshall's book, and what mention there is, is mostly incidental and for the most part relegated to footnotes or appendices. This is in marked contrast to the treatises of Smith, Ricardo, and Mill, in which issues of taxation (as well as public expenditure) occupied a major part of their presentation of the principles of economics. A possible explanation of this neglect on the part of Marshall is that he initially saw his *Principles* as the first of a work in two volumes, where the second volume was to contain the application of theory to several areas of economic policy; a sketch of the proposed contents of Volume 2 dated in October 1887 lists "Taxation" as one of six such areas, whereas in 1903, "Public finance" had become one of nine areas. When his *Industry and Trade* was finally published in 1919, these topics were no longer parts of the content of the book.[26]

[26] An interesting study of Marshall's plans for a second—and a third and possibly a fourth—volume of the *Principles* is Whitaker (1990).

1.3.2.3 J. B. Clark

John Bates Clark was a pioneer of the modern marginalist thinking in the United States who introduced the concepts of marginal productivity and marginal utility both in academic and more popular writings. But his 1899 book, *The Distribution of Wealth*, has become less known for its restatement of marginal productivity theory (which is its main focus) than for what Stigler (1941) referred to as its "naïve productivity ethics." In Clark's view, the equality between factor prices and marginal value productivity was not just a descriptive theory of how the market worked; it was also the manifestation of a natural law. This view is expressed already on the first page of the preface:

> *It is the purpose of this work to show that the distribution of the income of society is controlled by a natural law, and that this law, if it worked without friction, would give to every agent of production the amount of wealth which that agent creates.*
>
> **Clark (1899, p. v)**

This statement may be read simply as a characterization of factor market equilibrium under perfect competition, although it raises the issue of how an agent's marginal productivity can be identified with "what he creates." Clark maintained that this problem was less complex than many people thought, for it was essentially of the same nature as that which arose in a simple frontier society:

> *In particular, it is necessary to know that the primitive law which puts a man face to face with nature and makes him dependent on what he personally can make her yield to him is still, in essence, the law of the most complex economy.*
>
> **Clark (1899, p. 37)**

A further and crucial issue is whether the distribution that results from the operation of the law is just. On this point, there is a certain ambivalence in Clark's exposition. On the one hand, he said that this question lies outside his enquiry, "for it is a matter of pure ethics" (p. 8). On the other hand, he argued that what he creates belongs to the agent by right, and that nobody can complain if he is paid according to what he creates. The competitive distribution of income is therefore both fair and consistent with social stability, for if some agents are paid less than what they create,

> *there would be at the foundation of the social structure an explosive element which sooner or later would destroy it.*
>
> **Clark (1899, p. 9)**

Although most modern economists will no doubt find Clark's "productivity ethics" unconvincing, there are also elements in his thought that have been taken up by others. The most obvious parallel is the analysis by the philosopher Robert Nozick in his book *Anarchy, State, and Utopia* (Nozick, 1974). Nozick's basic idea is what he calls the entitlement theory of distributive justice. Any distribution that reflects an acquisition of income or wealth that is considered to be fair, i.e., to have been fairly acquired according to certain axiomatic criteria, is just. Moreover, given such a distribution, there is no case for public redistribution of income. Although it is not linked to the marginal productivity theory of income distribution, Nozick's theory evidently has some elements in common with the ideas of Clark.[27]

1.3.2.4 Pareto

We have already encountered Pareto as an empirical researcher on income distribution. Although his influence in that area was significant, his contribution to welfare economics was more fundamental and of more lasting significance. It had important consequences for the way that economists thought about normative issues, including their views on income redistribution as a goal of economic policy.

The starting point for Pareto's welfare economics was his study of utility and demand. Arguing in his *Manual of Political Economy* that only an ordinal concept of utility was required as a foundation for the study of consumers' demand,[28] he went on to point out that this concept of utility did not lend itself to interpersonal comparisons:

> *The utility, or its index, for one individual, and the utility, or its index, for another individual, are heterogeneous quantities. We can neither add them together nor compare them ... A sum of utility enjoyed by different individuals does not exist; it is an expression which has no meaning.*
>
> **Pareto (1909; 1971, p. 192)**

From this it would seem to follow that the search for a criterion of aggregate utility or welfare would be in vain. However, Pareto went on to introduce his own criterion of social welfare or efficiency that we now call Pareto optimality:

> *We will say that the members of a collectivity enjoy* maximum utility *in a certain position when it is impossible to find a way of moving from that position very slightly in such a manner that the utility enjoyed by each of the individuals of that collectivity increases.*
>
> **Pareto (1909; 1971, p. 261)**

"Maximum utility" was clearly not a good name for this concept because it suggested precisely the type of aggregation that Pareto sought to avoid, but he may be excused for not inventing the term "Pareto optimality."

Pareto showed that a competitive equilibrium satisfied the conditions for optimality in this sense. From the assumption of incomparability, it followed that his optimality criterion was unable to judge the welfare effects of a redistribution of income that led to diminished incomes for the rich and increased incomes for the poor because this would make the rich enjoy less utility and the poor more. If the economy were to find itself in a competitive equilibrium both before and after the redistribution of income, both states of

[27] Ideas similar to those of Nozick had also been advanced by Friedrich Hayek; see in particular Hayek (1973).

[28] To distinguish this concept from that of cardinal utility, he even coined a new word, *ophelimity* (*ophélimité*), to represent it—a word that never caught on. In the quotations that follow, I have substituted "utility" for Pareto's "ophelimity."

the economy would satisfy the conditions for Pareto optimality, but the optimality criterion would not be able to rank the two situations relative to each other. Judgments about income distribution and redistribution in terms of justice or fairness should, according to this view, be regarded as occupying a position outside the field of economics as a scientific discipline. Although this interpretation is not very explicit in Pareto's own work, it became a central proposition in the further elaboration of Paretian welfare economics that was carried out by a number of twentieth-century economists. But the acceptance of Pareto optimality as an important concept of welfare economics took a long time. As late as 1947 Paul Samuelson, after having presented the definition of Pareto optimality, could write that "it has not yet received attention from economists commensurate with the importance which he [Pareto] attached to it" (Samuelson, 1947, p. 212).

1.3.3 Utilitarianism and the Economics of Redistribution

The insistence by Walras and even more strongly by Pareto on the subjective nature of utility might have been expected to lead to the total banishment of utilitarian philosophy from the normative analysis of income distribution. However, this did not happen. There were several reasons for this. One is that the work of Walras and especially Pareto did not become widely known in the international community of economists until well into the twentieth century. Another was that utilitarianism continued to hold a strong attraction for economists in search of a philosophical foundation for their egalitarian convictions and for the design of redistributive policy, particularly in the tax field.

1.3.3.1 Maximizing the Sum of Utilities

A good example of such an economist is Francis Ysidro Edgeworth. He adopted the view of the older utilitarians that social welfare should be seen as the sum of individual utilities but was critical of the use that they made of it, pointing out that it was difficult to see, in the absence of mathematical formalization, how their conclusions followed from their ethical premises. In his book *New and Old Methods of Ethics* (Edgeworth, 1877) he built on the analogy with the Weber–Fechner law in psychology, which stated that the perception of a sensual stimulus increases less than proportionally with the strength of the stimulus, to argue that utility must increase less than proportionally with income. From this he drew strong conclusions for the socially optimal distribution of income. In the case of a given total income to be divided between all members of society the optimal distribution would be one of complete equality, assuming that all individuals had the same utility function of income. He also analyzed the case of variable work effort and found that under certain assumptions those with the greatest capacity should do the most work.

A related approach was that of Pigou. In his *Economics of Welfare* (1920), he used an explicit utilitarian argument—although without reference to the Weber–Fechner law—to argue in favor of redistribution of income from the rich to the poor:

> *... it is evident that any transference of income from a relatively rich man to a relatively poor man of similar temperament, since it enables more intense wants to be satisfied at the expense of less intense wants, must increase the aggregate sum of satisfaction. The old "law of diminishing [marginal] utility" thus leads securely to the proposition: Any cause which increases the absolute share of real income in the hands of the poor, provided that it does not lead to a contraction in the size of the national dividend from any point of view, will, in general, increase economic welfare.*
>
> **Pigou (1920; 1932, p. 89)**

In other words, it is assumed that there exists a utility function of income that is concave and the same for everybody. In the following pages, the proviso of "similar temperament" is spelled out further. Pigou admits that under existing social conditions a rich man may in fact be able to produce more utility from any given amount of income than a poor man. But this advantage has come about through past inequalities of income and the standard of living and cannot therefore be used to argue against income equalization: In the long run, the poor who experience increased incomes will be as able as the current rich to generate utility from their income. The last part of the quotation introduces an important qualification: Policies that aim to redistribute income from the rich to the poor may have an adverse effect on incentives, in particular on the incentives to work and save. This may lead to a reduction of the national dividend or national income so that there will be less income available for distribution.

In analytical terms, we might restate this argument as saying that if there are no incentive effects of redistribution it should be carried to the point where the marginal utility of income is the same for all; in the case of identical tastes, this would imply complete equalization of incomes, as in the analysis of Edgeworth. If incentive effects are present, the optimal amount of redistribution would stop short of this point, with the gap between the marginal utility of income between rich and poor determined by the strength of the incentive effects.

1.3.3.2 Critique of Utilitarianism

The assumptions of identical utility functions, decreasing marginal utility, and interpersonal comparability of utility all became the subject of critical scrutiny as Pareto's work on demand theory and welfare economics became more widely known. Because these assumptions had been shown to be unnecessary for the study of consumer demand, they were also held to be inappropriate for making welfare judgments. Justifications of income redistribution such as that advanced by Pigou gradually came to be viewed as nonscientific and simply subjective expressions of one's personal taste for income equality. On the desirability of redistribution, economics as a science would have to remain silent. This view was particularly forcefully put in the influential book by Lionel Robbins (1932).

Robbins's influence is clearly discernible in the New Welfare Economics that was developed by several writers during the 1930s and 1940s. In the reformulation of welfare theory by Bergson (1938) and Samuelson (1947), a crucial role was played by the social welfare function that depicted social welfare as an increasing function of individual utility levels, represented by ordinal utility functions. The conditions for social welfare maximization could then be stated as two set of conditions. One set described the conditions for Pareto optimal allocation of factors of production and consumer goods, whereas the other represented the conditions for optimal distribution between consumers as requiring equality of the social marginal utility of income—i.e., the increase in social welfare following an increase in income—between individuals.[29] Although the new formulation made clear the distinction between welfare judgments related to efficiency on the one hand and distributive justice on the other, the generality of the conditions that Samuelson (1947) referred to as the interpersonal optimal conditions was such that it became virtually impossible to draw any conclusion regarding the socially desirable form of income redistribution. At the most general level of analysis, the only conclusion that could be drawn from the analysis was that the desirable extent of redistribution was determined by one's ethical beliefs. Regarding the *form* of redistribution, however, the analysis had rather strong implications: To achieve a full optimum of social welfare, redistribution ought to be carried out by means of instruments that did not lead to violation of the efficiency conditions. The only instruments that could achieve this were individualized lump sum taxes and transfers (although some economists, e.g., Hotelling (1938), implicitly assumed that the income tax was at least approximately equivalent to lump sum taxation).

1.3.3.3 A Comeback for Utilitarianism

Although the new welfare economics helped to clarify the relationship between economists' statements regarding efficiency and distributive justice, one might still ask whether the representatives of the new approach went too far in their rejection of the old welfare economics, which was based on a cardinal definition of utility and interpersonal utility comparisons. This view has been argued by Cooter and Rappoport (1984), who maintained that the concepts of utility used by the post-Pareto ordinalist school and the older economists whom they refer to as the material welfare school were fundamentally different. The concept of utility employed by the material welfare school was not intended to represent the individual's tastes but his needs, and these needs were assumed to be objectively observable as for instance in the form of physical fitness. To use this concept for interpersonal comparisons did not involve a comparison of subjective preferences but of empirically observable standards of living. The consumption goods that were bought using the individual's income were used to produce his standard of living, but like other factors of production the goods obeyed the law of diminishing returns,

[29] The implication is that Pareto optimality is a necessary but not sufficient condition for a maximum of the social welfare function. This follows directly from the assumption that the social welfare function is increasing in individual utilities.

which in this case was translated into the concept of diminishing marginal utility of income. It was this concept of utility that was used by economists like Edgeworth[30] and Pigou to justify the recommendation of transfers to the poor and progressive taxation. The concreteness of the concept is well brought out in Hugh Dalton's (1920) comment on Jevons's (1871) discovery[31] of the law of diminishing marginal utility.

> *From this law a practical conclusion of the greatest importance follows, namely, the extreme wastefulness from the point of view of economic welfare of large inequalities of income. It is obvious to the modern economist that, from this point of view, a considerable equalization of incomes is desirable, provided that production is not checked thereby. But before Jevons wrote, this was by no means obvious, or at any rate it was not widely perceived.*
>
> **Dalton (1920, p. 90)**

Dalton's use of the word *wastefulness* is suggestive. In the new welfare economics framework, this term would be meaningless, but in the approach taken by the material welfare school, it has a concrete interpretation in terms of a smaller quantity of aggregate welfare, which is due to the inequality of income. Given the way that income is distributed, it produces a smaller amount of material welfare or standard of living than that which would result from a more equal distribution.

A new justification for the utilitarian social welfare function arose in the early postwar period. It started with an article by William Vickrey (1945), which was apparently concerned with the possibility of measuring the marginal utility of income on the basis of the von Neumann–Morgenstern expected utility hypothesis. But in the middle of the article, Vickrey changed his focus to that of discussing the question of the socially optimal distribution of income. His approach is nicely summed up in the following statement:

> *If utility is defined as that quantity the mathematical expectation of which is maximized by an individual making choices involving risk, then to maximize the aggregate of such utility over the population is equivalent to choosing that distribution of income which such an individual would select were he asked which of various variants of the economy he would like to become a member of, assuming that once he selects a given economy with a given distribution of income he has an equal chance of landing in the shoes of each member of it.*
>
> **Vickrey (1945, p. 329)**

The idea was developed further by several writers, including Marcus Fleming (1952) and John Harsanyi (1955b), neither of whom, however, referred to Vickrey's work. Harsanyi's article in particular showed how a utilitarian social welfare function, additive in individual utilities, could be derived from a set of axioms governing individual and social welfare judgments. Using this approach, one could go back to the issue raised by the earlier utilitarian economists and ask which distribution of a given amount of

[30] Samuelson (1947, p. 206) said that "to a man like Edgeworth, steeped as he was in the Utilitarian tradition, individual utility—nay social utility—was as real as his morning jam."

[31] Jevons was in fact not the first to formulate this principle. As he was later to acknowledge, Gossen (1854) had done so before him. Even earlier, although in a different context, the principle had been formulated by Bernoulli (1738). For references and further discussion, see Sandmo (2011).

income would maximize social welfare. If social welfare can be expressed as an unweighted sum of individual utility functions, and if these functions are concave (representing risk-averse attitudes), the answer would once again be that the optimal distribution would be one of complete equality.

This implication was not emphasized by Harsanyi, whose interests centered on the logical foundations for this particular social welfare function, not in its implications for social organization and economic policy. Vickrey, on the other hand, developed these implications in some detail, pointing out both the optimality of equal distribution if total income could be taken as fixed and the qualifications needed when one takes account of the objection that the total amount of income cannot in practice be taken as independent of the way it is distributed.[32] Therefore, he argued, "some degree of inequality is needed in order to provide the required incentives and stimuli to efficient cooperation of individuals in the production process" (Vickrey, 1945, p. 329). From this observation, he proceeded to an attempt to determine the welfare maximizing amount of redistribution by calculating an optimal income tax function using the calculus of variations. He succeeded in deriving the Euler equation for this problem but concluded that "even in this simplified form the problem resists any facile solution" (Vickrey, 1945, p. 331).

There is a direct line from Vickrey's analysis to the modern theory of optimal income taxation as pioneered by James Mirrlees (1971). Mirrlees also adopted the utilitarian assumption of social welfare as the sum of individual utility functions (which he also assumed to be identical) but without the choice theoretic foundation adopted by Vickrey and Harsanyi; it is also notable that he does not refer to Vickrey's (1945) article. In the Mirrlees model, individual utility functions depend on consumption (or income) and leisure. Lump sum taxation is ruled out as infeasible, and redistribution has to be carried out by means of a nonlinear income tax that distorts the choice between leisure and consumption. The shape of the optimal income tax function accordingly has to reflect the trade-off between equality and efficiency. By adopting some additional assumptions relative to Vickrey's model, Mirrlees was in fact able to characterize the optimal income tax function, although in rather general terms. More specific results were derived by a simulation analysis of special cases. A surprising feature of the optimal tax schedule that emerged from these numerical experiments was that although the average tax rate was increasing in income, the marginal tax rate tended to stay approximately constant and if anything showed a tendency to decline with income.[33] Mirrlees's contribution has led to a long line of refinements and extensions of his analysis, including a critical examination of the utilitarian foundations of the social welfare function. In the 1970s, the book

[32] Vickrey's argument is strongly reminiscent of that of Pigou in the *Economics of Welfare*, but he made no reference to Pigou or any of the other early utilitarian economists.

[33] Later work by several economists demonstrated that under certain assumptions the optimal marginal rate of income tax at the top of the income schedule should be equal to zero. For an interpretation of this result and further references, see, e.g., Sandmo (1999).

by the philosopher John Rawls (1972) created a great deal of interest among economists who were interested in public policy analysis, and Rawls's "maxi-min" criterion, by which the welfare criterion to be maximized is the utility of the least fortunate person in society, was applied to the problem of optimal income taxation by Atkinson (1973). His numerical results indicated that with this criterion the marginal tax rates and the degree of progression were likely to be considerably higher than in the case considered by Mirrlees.

As an aside, it may be noted that a different argument for low marginal tax rates had earlier been discussed by Ragnar Frisch in an article published in Norwegian (Frisch, 1948). Frisch based his argument on the distinction between what he called the internal and external marginal productivity of labor. The external marginal productivity in a particular sector refers to the effect on output in other sectors that is not taken into account in the employment decision. Frisch believed that this effect as a rule was positive, so that work effort tended to be too low in a market economy. This might call for a negative marginal tax rate,[34] which, however, was not practically feasible, "at least not at the present time." Instead, he suggested a zero marginal tax rate on the part of income that was directly related to effort, and the remainder of the individual's income could be taxed according to a progressive scale.

1.3.4 Sacrifice and Benefit Theories

There are other ways to analyze the normative problems of redistribution than via social welfare maximization, and in this section, we consider two of these. Equal sacrifice theories caught the attention of economists around the end of the nineteenth century and were for a time influential in policy debates. Benefit theories of taxation whereby taxes are seen as payment for benefits received from the state have traditionally had a strong appeal to those who look for fairness in the relationship between the individual and the state.

1.3.4.1 Equal Sacrifice

The utilitarian approach to income distribution and taxation is sometimes referred to as an equal sacrifice theory. In the simple case that forms the starting point for the utilitarian analysis, pretax incomes are given and the government aims to collect a given amount of revenue by using individualized lump sum taxes to maximize the sum of identical and concave utility functions of income. The resulting optimal distribution of after-tax incomes is one of complete equality of income where the marginal utility of income is the same for all. The solution represent a minimum of aggregate sacrifice because the outcome with equal marginal utilities of income is the maximum of total utility that can be obtained relative to the tax revenue that is to be collected. It is a solution of equal sacrifice between persons only in the sense of equal *marginal* sacrifice: The sacrifice of the last dollar paid in taxes is the same for all.

[34] Or in other words a Pigouvian subsidy to work effort, although Frisch does not use this terminology.

It might be expected that some economists who thought about the just distribution of the tax burden should come to think that this notion of equal sacrifice had limited appeal. In the case of substantial inequality of pretax incomes, the loss of utility from going from the pretax to the after-tax situation will obviously differ between individuals, and if one thinks that this is unjust it is natural to look for some alternative notion of equal sacrifice that could be applied to such nonmarginal changes in the distribution of income. This led to the development of equal sacrifice theories in the more specific sense, and in particular the theories of equal absolute and equal proportional sacrifice; theories that were first discussed analytically by Cohen-Stuart (1889) and Edgeworth (1897). The criterion of equal absolute sacrifice)[35] can be formalized as:

$$U(Y) - U(Y - T) = k.$$

Here, Y is pretax income and T is the amount of tax, while k is a constant that is the same for all taxpayers,[36] so that the sacrifice of utility that results from taxation is the same for all individuals. To see how the amount of tax varies with income according to this principle, one may take the derivative of the left-hand side of the equation with respect to Y, treating T as a function of Y. Solving for the marginal tax rate, we obtain:

$$\mathrm{d}T/\mathrm{d}Y = [U'(Y - T) - U'(Y)]/U'(Y - T).$$

One sees immediately that the assumption of decreasing marginal utility of income implies that the marginal tax rate is positive, but the assumption does not take us any further in supplying an argument for progressive taxation. To study the implications for progressivity, one can use the result to derive the elasticity of income after tax with respect to income before tax. For progressivity, this should be less than one, but whether this is the case or not turns out to depend on whether the elasticity of the marginal utility of income is less than or greater than -1. For the logarithmic function, where the elasticity is just -1, equal sacrifice in this sense implies proportional rather than progressive taxation, as pointed out by Samuelson (1947, p. 227). From the point of view of the history of public finance, this conclusion is of particular interest because it was for some time widely believed that the principle of equal sacrifice combined with the assumption of decreasing marginal utility of income was sufficient to justify progressive taxation.[37]

35 The central contributions of Cohen-Stuart and Edgeworth have been reprinted in Musgrave and Peacock (1958). The criterion of equal relative or proportional sacrifice, whereby the difference in utility levels is related to the before tax utility level, leads to slightly different conditions for progressive taxation but does not raise any new issues of principle. See Musgrave (1959, p. 96).

36 k must reflect the government's revenue requirement, so that T is higher, the higher is k.

37 Cohen-Stuart (1889; 1958) surveyed a number of earlier studies of this issue by German and Dutch writers who claimed that progressive taxation could be rationalized along these lines. See also the book by Blum and Kalven (1953), which surveys both economic and legal discussions of tax progressivity with emphasis on the arguments derived from equal sacrifice theories.

Although the principle of equal sacrifice may have some appeal to economic intuition, the main reason that it has disappeared from the modern discussion of optimal redistribution must be that its assumptions are difficult to reconcile with the maximization of a social welfare function. From that perspective, the straightforward utilitarian approach is much more appealing. In addition, the equal sacrifice theory lends itself less easily to generalizations incorporating variable labor supply and the second best considerations introduced by the work of Mirrlees and others into the utilitarian framework. From this point of view, the equal sacrifice theory of income redistribution proved to be a sidetrack.[38]

1.3.4.2 The Benefit Principle of Taxation

The utilitarian and related approaches to the issue of optimal income distribution considered the question of the just or fair distribution of income in isolation from the distributive effects of public expenditure. In the older literature, we have seen that Adam Smith recommended that the contributions of taxpayers should be in proportion to "the revenue which they respectively enjoy under the protection of the state," and one interpretation of this rule is that taxes should be levied so as to correspond to the benefits that people received from the activities of the state. However, the further elaboration of the benefit principle of taxation mainly took place in the writings of a number of continental European economists during the late nineteenth and early twentieth centuries. Two different types of claims were made for the implementation of the benefit principle of taxation. The first was that taxes levied on individuals according to the benefits that they received from the provision of public goods would somehow establish a price system for public goods or publicly provided goods that would correspond to competitive prices for private goods with similar efficiency properties. This idea suffers from the weakness that at least for public goods in the proper sense these prices do not provide individuals with the incentives to reveal their true preferences so that they cannot fill the functions of the price mechanism in the private goods part of the economy. The second claim, which is the one that is relevant for the normative analysis of redistribution policy, is that the benefit principle represents justice in taxation and that it therefore is important for normative judgments about income distribution in a mixed economy. The best known statement of this position is that of Knut Wicksell (1896).[39]

The concept of just taxation as used by Wicksell is quite different from that employed by economists in the utilitarian tradition. Wicksell sees the relationship between government and citizens as basically one of exchange, and one that should be carried out on

[38] Or, as put by Edgeworth (1897, p. 566): ". . . whatever view we take of the relation of the principle of like sacrifice to pure utilitarianism, the sphere of its action independently of that supreme principle appears to be insignificant."

[39] The collection of translations edited by Musgrave and Peacock (1958) contains many of the most important contributions to this line of analysis by German, Italian, and Swedish economists, including a central extract from Wicksell's book.

terms that are fair. The starting point for his argument is that no public project should be carried out unless society's aggregate willingness to pay is at least as high as its costs. Given that this condition is satisfied, it ought to be possible to distribute the costs in such a manner that every citizen makes a gain from the exchange, and this is the principle of justice in taxation: "No-one can complain if he secures a benefit which he himself considers to be (greater or at least) as great as the price he has to pay" (Wicksell, 1896; 1958, p. 79). From this he drew the conclusion that any political proposal about public projects should be voted on as a balanced budget tax-expenditure "package," and that it should only be passed on the basis of a unanimous vote.

It may seem surprising that Wicksell with his reputation for political radicalism should favor a system that seems to exclude the possibility of income redistribution through the public budget. It is at this point that one has to keep his peculiar definition of "just taxation" in mind. Wicksell said explicitly that the principle does not take account of distributional issues. Given the distribution of income in society, Wicksell's principle, as described here, does nothing more than assure that the adoption of any new public project does not harm any citizen.[40] He also emphasized that this principle, if adopted in the Swedish society of his own time, would be in the interests of the lower classes who in his view were exploited by the higher income groups to contribute to the financing of public projects that involved little or no benefit to themselves.

However, Wicksell recognized that for this principle to be fully convincing both from an economic and ethical point of view, it would have to be embedded in a broader framework of distributive justice: "It is clear that justice in taxation tacitly presupposes justice in the existing distribution of property and income" (Wicksell, 1896; 1958, p. 108). On this broader concept of justice, however, he has actually little to say, although he emphasized that too much redistribution may harm the upper classes in a way that is harmful to society as a whole because these classes "undeniably include a significant share of a nation's intelligence and economic initiative" (Wicksell, 1896; 1958, p. 117).

Wicksell's analysis was followed up by his countryman Erik Lindahl, whose monograph on the theory of taxation introduced the concept that later came to be known as Lindahl prices (Lindahl, 1919). In a later article, he discussed in more detail the argument that the benefit principle had a claim to be considered a standard of justice in taxation. Here, on the one hand, he emphasized the broader concept of distributive justice in which the benefit principle had to be embedded[41]:

[40] The principle is closely related to the Pareto principle, which identifies a social improvement with the case where no one becomes worse off. Wicksell later modified the unanimity requirement to apply to groups rather than individuals.

[41] The term *property* should here be interpreted in a broad sense as including all individual economic resources, including income.

> *. . . justice in taxation is inextricably linked with justice in the distribution of property, since it would obviously be nonsense to speak of "a just portion of an unjust whole."*
>
> **Lindahl (1928; 1958, p. 227)**

On the other hand, Lindahl also argued that there did not necessarily exist any contradiction between the principles of benefit and ability to pay because ability to pay could often be taken as a good indication of the benefit derived from public expenditure. On this point, Lindahl's argument is reminiscent of Adam Smith's first maxim of taxation, which indicated that it would be possible for taxation simultaneously to reflect both the individual taxpayers' ability to pay and the benefits that they received under the protection of the state.

1.4. CONCLUDING REFLECTIONS

A chapter of the history of economic thought regarding income distribution theories does not lend itself easily to a summary in the way of a few main conclusions. Looking back on a two-hundred-year-long history, however, it does induce one to offer a few general reflections on the nature of the field and its development. Accordingly, I will make a few remarks on two general issues. The first concerns the relationship of economic theory to empirical evidence, in particular before the time when econometrics was established as the main framework for empirical study. The second set of remarks relates to Ricardo's characterization of income distribution as the principal problem of economics: Does the history of economic thought confirm his view of the importance of the subject?

1.4.1 Theory and Evidence

This chapter has been primarily an account of *theories* of income distribution; to include also the statistical and empirical work that has been done over the two centuries covered by the survey would be impossible within the confines of a single article. However, a brief discussion may be in order regarding the connection between theoretical and empirical work during the period. Thus, an interesting question to consider is to what extent the theorists of income distribution were aware of and were influenced by the empirical work that was undertaken at about the same time. In particular, the nineteenth century witnessed the growth of official statistics covering both the development of national income and its distribution.

The questions of awareness and influence are very general, and it is not easy to provide clear and simple answers. One reason for this is that the influence of empirical knowledge on economic theorists may have been rather indirect; some characteristics of the real economy may have been considered to be common knowledge, so that theorists saw no need to provide exact documentation. But one should realize that there was not always agreement about what that supposedly common knowledge actually was. A case in point is John Stuart Mill's disagreement with Adam Smith regarding the structure of wages. As we have seen, Smith believed that labor market competition would ensure that occupational wage rates would tend to compensate for noneconomic advantages and disadvantages, whereas Mill claimed that quite to the contrary, wage differentials reinforced the inequalities arising from different working conditions. In the eighteenth century, empirical data on this issue were presumably hard to come by; nevertheless, Smith did refer to empirical observations in support of his hypothesis, although by modern standards these references are both incomplete and unsystematic. By the middle of the next century, however, the situation had changed, and it would have been possible for Mill to provide if not direct evidence at least some empirical illustrations that could throw light on this matter and more generally on the distribution of income. But he obviously felt no need to do this. Well into the next century, Hicks (1932) wrote about the effects of competition on the structure of wages with hardly any reference to empirical relationships. In fact, the only instance in which he did refer to empirical evidence is where he cited data for wages of agricultural laborers in Lancashire in 1794, showing how they vary with the distance to the nearest manufacturing center.[42]

To blame the economists of the eighteenth and nineteenth centuries for not supplying formal statistical tests of their theories would of course be pointless because at that time econometrics was not even in its infancy (see Morgan, 1990). What one might nevertheless have expected was a greater interest in drawing on data that could illustrate the importance and relevance of theoretical reasoning.

From this point of view, a more striking instance of the lack of connection between theoretical and empirical work is Knut Wicksell's belief that real wages had not risen significantly over the past two centuries, as seen from the perspective of 1901. At that time, there had actually accumulated a large amount of statistical data documenting the significant rise in real wages during the nineteenth century in countries such as Germany, Italy, Great Britain, and the United States (see, e.g., Bresciani-Turroni, 1939) as well as the Scandinavian countries. The data for Germany were particularly extensive and at Wicksell's time had been used in academic studies by several German economists. Wicksell read (and wrote) German, but for whatever reason this work had little or no influence on his own thinking. If he had utilized it, he would have seen that his belief was firmly rejected by the empirical evidence.[43] But at least on this particular issue, he must have felt no need to confront his theoretical conclusions with statistical evidence.[44]

[42] These data were drawn from Redford (1926).

[43] In the case of Sweden, later economic historians have found that real wages increased at an annual rate of more than 2% during the period 1860–1895 (Phelps Brown and Browne, 1968). Although these particular statistics were not available to Wicksell, it is hard to imagine that this growth was not noticeable for people living at the time.

[44] Wicksell must also have known several economists and statisticians who had personal experience of the empirical work. Thus, he had extensive contacts with Norwegian economists, among whom was Anders Nicolai Kiær, who was an acknowledged expert on income and wealth statistics.

The history of the interaction—or lack of it—between theorists and empirical researchers in the study of income distribution is a large topic in itself that cannot be surveyed here, particularly because it cannot be separated from the broader issue of the connection between theoretical and empirical research in economics more generally. The present examples of the lack of such a connection should simply be taken as an indication that at least in the preeconometric age, there were sometimes large gaps between theoretical and empirical insights.

1.4.2 The Principal Problem of Political Economy?

Toward the end of this review of the development of theories of income distribution, it is natural to reconsider the quotation from Ricardo with which we began. Is Ricardo's view reflected in the actual importance that the theory of income distribution has had in the history of economics? It may well have been true that Ricardo in this way expressed his conviction of the nature of economics, but his vision must be interpreted in light of the state of the science at the time in which he lived as well as the nature of society. To a modern economist, the proposition that the functional distribution of income between workers, capitalists, and landowners should be considered the most important problem in economics will hardly be a convincing one. One of the reasons why Ricardo gave such emphatic priority to the problem may have been a conviction that the analysis of this issue also went far to explain the personal distribution in a society with a modest degree of mobility between social and economic classes. Another reason may have been that he did not see the economic theory of his day as providing a set of analytical tools and concepts that would be useful in a more disaggregated analysis of the personal distribution of income.

All this has of course changed. With the marginalist revolution of the late nineteenth century, economists acquired a set of theoretical tools that gradually came to improve their opportunities for analysis of both the positive and normative aspects of income distribution. But to what extent did they exploit these opportunities? When one reads the contributions of the early marginalists, it becomes obvious that they applied their new theories mainly to the explanation of price formation in the market for commodities and less so in the markets for factors of production. When the general equilibrium followers of Walras put the finishing touches to the neoclassical theory of competitive markets, commodities and factors were treated symmetrically with the result that less attention was given to the special features of the markets for labor, capital, and natural resources. Labor economics was for a long time considered to be a field on the outskirts of theory-based economics, and the literature on financial markets paid little or no attention—and continues to pay little or no attention—to the study of the personal distribution of income and wealth. Only in recent decades has formal economic theory begun to catch up on its neglect of the determination of income distribution. But this neglect is still visible in the allocation of space in introductory textbooks and books on microeconomic theory.

These remarks pertain in particular to the positive economics of income distribution. But the attention to normative issues has fluctuated even more. Questions of distributive justice were certainly discussed by the classical economists but without the benefit of a formal theoretical structure. With the breakthrough of marginal utility theory the situation changed, and many economists saw no objection to utilizing the hypothesis of decreasing marginal utility both to explain consumer demand and to justify the utilitarian argument in favor of income equality. This approach suffered a setback in the early nineteenth century with the adoption of ordinalism and the ideas of a value-free science. Later on, it once again became accepted that welfare economics could make an important contribution in clarifying the borderline between statements of facts and values, whereas since the 1960s, as Atkinson (2001) has pointed out, many modern textbooks seem to have adopted the view that the basic elements of welfare economics do not form a central part of the training of the modern economist. In regard to the theory of income distribution, many economists seem be held back from a discussion of distributive justice, presumably because it will lead them into areas where they have to confront issues that are of an ethical or philosophical nature.[45]

The desirable awareness of the relationship between positive and normative approaches to issues of income distribution may also be promoted through better knowledge of the history of thought in the area. Here, there is definitely room for improvement. History provides many examples of how new theories have been formulated without apparent awareness of the work of earlier economists. As an example, the modern theory of optimal income taxation could probably have been developed and presented with a broader appeal to the general economics profession if it had been set in the context of the work by earlier utilitarian economists such as Edgeworth and Pigou. It is undeniable that economics has many of the features of a cumulative science in which new theories replace old ones because of their higher explanatory power or because they lead to better insights in the problems that arise in the design of economic policy. But even a cumulative science can benefit from awareness of its roots.

ACKNOWLEDGMENTS

I am indebted to Tony Atkinson, Branko Milanovic, Martin Ravallion, and the participants in the authors' conference in Paris for many helpful comments and suggestions.

[45] A reviewer of my book on the history of economic thought (Sandmo, 2011) wrote, "I hate the word *social justice* because I do not know what it means."

REFERENCES

Arrow, K.J., Hahn, F.H., 1971. General Competitive Analysis. Holden-Day, San Francisco.

Asimakopulos, A. (Ed.), 1987. Theories of Income Distribution. Kluwer Academic Publishers, Boston.

Atkinson, A.B., 1970. On the measurement of inequality. J. Econ. Theory 2, 244–263.

Atkinson, A.B., 1973. How progressive should income tax be? In: Parkin, M., Nobay, A.R. (Eds.), Essays in Modern Economics. Longman, London, pp. 90–109, Reprinted in: Atkinson, A.B. Social Justice and Public Policy. Wheatsheaf Books, Brighton, 1983, pp. 295–314.

Atkinson, A.B., 1975. The Economics of Inequality. Oxford University Press, Oxford.

Atkinson, A.B., 2001. The strange disappearance of welfare economics. Kyklos 54, 193–206.

Atkinson, A.B., Bourguignon, F., 2000. Income distribution and economics. In: Atkinson, A.B., Bourguignon, F. (Eds.), Handbook of Income Distribution, vol. 1. Elsevier North-Holland, Amsterdam, pp. 1–58.

Becker, G.S., 1957. The Economics of Discrimination. University of Chicago Press, Chicago, second ed., 1971.

Becker, G.S., 1962. Investment in human capital: a theoretical analysis. J. Polit. Econ. 70 (5, part 2), 9–49.

Becker, G.S., 1964. Human Capital. Columbia University Press, New York, second ed., 1975.

Becker, G.S., Chiswick, B.R., 1966. Education and the distribution of earnings. Am. Econ. Rev. 56, 358–369.

Becker, G.S., Tomes, N., 1979. An equilibrium theory of the distribution of income and intergenerational mobility. J. Polit. Econ. 87, 1153–1189.

Bergson, A., 1938. A reformulation of certain aspects of welfare economics. Q. J. Econ. 52, 310–334.

Bernoulli, D., 1738. Specimen theoriae novae de mensura sortis. Commentarii Academiae Scientarium Imperiales Petropolitanae, Translated as: Exposition of a new theory on the measurement of risk. Econometrica 22, 23–36 (1954).

Blum, W.J., Kalven Jr., H., 1953. The Uneasy Case for Progressive Taxation. University of Chicago Press, Chicago.

Böhm-Bawerk, E.v., 1884-1889. Kapital und Kapitalzins. Fischer, Jena, Translated as: Capital and Interest. Libertarian Press, South Holland, IL, 1959.

Bresciani-Turroni, C., 1939. Annual survey of statistical data: Pareto's Law and the index of inequality of incomes. Econometrica 7, 107–133.

Cairnes, J.E., 1874. Some Leading Principles of Political Economy Newly Expounded. Macmillan, London.

Cannan, E., 1893. A History of the Theories of Production and Distribution in English Political Economy from 1776 to 1848, third ed., 1917. Reprint: Routledge/Thoemmes Press, London, 1997.

Cannan, E., 1914. Wealth. A Brief Explanation of the Causes of Economic Welfare, third ed., 1928. Reprint: Routledge/Thoemmes Press, London, 1997.

Cantillon, R., 1755. Essai sur la nature de la commerce en général. Translated as: Essay on the Nature of Commerce. Macmillan, London, 1931.

Chamberlin, E.H., 1933. The Theory of Monopolistic Competition. Harvard University Press, Cambridge, MA.

Champernowne, D.G., 1953. A model of income distribution. Econ. J. 63, 318–351.

Chipman, J.S., 1974. The welfare ranking of Pareto distributions. J. Econ. Theory 9, 275–282.

Chipman, J.S., 1976. The Paretian heritage. Revue Européenne des Sciences Sociales 14 (37), 65–173.

Clark, J.B., 1899. The Distribution of Wealth. A Theory of Wages, Interest and Profits. Macmillan, New York.

Cohen-Stuart, A.J., 1889. Bijdrage Tot De Theorie Der Progressieve Inkomstenbelasting. Martinus Nijhoff, The Hague, Translated in part as: On progressive taxation. In Musgrave, R.A., Peacock, Alan T. (Eds.), 1958. Classics in the Theory of Public Finance. Macmillan, London, pp. 48–71.

Cooter, R., Rappoport, P., 1984. Were the ordinalists wrong about welfare economics? J. Econ. Lit. 22, 507–530.

Dalton, H., 1920. Some Aspects of the Inequality of Incomes in Modern Communities. Routledge, London.

Debreu, G., 1959. Theory of Value. Wiley, New York.

Domar, E.D., Musgrave, R.A., 1944. Proportional income taxation and risk-taking. Q. J. Econ. 58, 388–422.

Dunlop, J.T., 1944. Wage Determination under Trade Unions. Macmillan, New York.

Dunlop, J.T., 1958. Industrial Relations Systems, second revised ed., Harvard Business School Press, 1993.

Edgeworth, F.Y., 1877. New and Old Methods of Ethics. James Parker, Oxford.

Edgeworth, F.Y., 1896. Supplementary notes on statistics. J. R. Stat. Soc. 59, 529–539.

Edgeworth, F.Y., 1897. The pure theory of taxation III. Econ. J. 7, 550–571.

Edgeworth, F.Y., 1926. Pareto's Law. Palgrave's Dictionary of Political Economy, vol. III, Appendix, 712–713. Reprinted in: Newman, P. (Ed.), F.Y.Edgeworth: Mathematical Psychics and Further Papers on Political Economy. Oxford University Press, Oxford, 2003, pp. 491–494.

Engel, E., 1883. Der Wert des Menschen. 1. Theil: Der Kostenwerth des Menschen. Verlag von Leonhard Simion, Berlin.

Engels, F., 1845. Die Lage der arbeitenden Klassen in England, Translated as: The Condition of the Working Class in England. Oxford University Press, Oxford, 1993.

Fleming, M., 1952. A cardinal concept of welfare. Q. J. Econ. 66, 366–384.

Friedman, M., 1953. Choice, chance, and the personal distribution of income. J. Polit. Econ. 61, 277–290.

Friedman, M., Kuznets, S., 1945. Income from Independent Professional Practice. National Bureau of Economic Research, New York.

Friedman, M., Savage, L.J., 1948. The utility analysis of choices involving risk. J. Polit. Econ. 56, 279–304.

Frisch, R., 1948. Den optimale arbeidsinnsats (The optimal work effort). Ekonomisk Tidskrift 50, 63–80.

Galbraith, J.K., 1958. The Affluent Society. Houghton Mifflin, Boston.

Goldfarb, R.S., Leonard, T.C., 2005. Inequality of what among whom? Rival conceptions of distribution in the 20th century. Research in the History of Economic Thought and Methodology 23-A, 75–118.

Gossen, H.H., 1854. Entwickelung der Gesetze des menschlichen Verkehrs und der daraus fliessenden Regeln für menschliches Handeln. Friedrich Vieweg & Sohn, Braunschweig, Translated as: The Laws of Human Relations and the Rules of Human Action Derived Therefrom. MIT Press, Cambridge, MA, 1983.

Harberger, A.C., 1962. The incidence of the corporation income tax. J. Polit. Econ. 70, 215–240.

Harsanyi, J.C., 1955a. Approaches to the bargaining problem before and after the theory of games: a critical discussion of Zeuthen's, Hicks's and Nash's theories. Econometrica 24, 144–157.

Harsanyi, J.C., 1955b. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. J. Polit. Econ. 63, 309–321.

Hayek, F.A., 1973. Law, Legislation and Liberty. University of Chicago Press, Chicago.

Hicks, J.R., 1932. The Theory of Wages. Macmillan, London.

Hotelling, H., 1938. The general welfare in relation to problems of taxation and of railway and utility rates. Econometrica 6, 242–269.

Jaffé, W. (Ed.), 1965. Correspondence of Léon Walras and Related Papers, vols. I–III. North-Holland, Amsterdam.

Jevons, W.S., 1871. The Theory of Political Economy. Macmillan, London, Pelican Classics Edition, edited by R. D. Collison Black. Penguin Books, Harmondsworth, 1970.

Kanbur, S.M., 1979. Of risk taking and the personal distribution of incomes. J. Polit. Econ. 87, 769–797.

Kanbur, S.M., 1981. Risk taking and taxation: an alternative perspective. J. Public Econ. 15, 163–184.

Kuznets, S., 1955. Economic growth and income inequality. Am. Econ. Rev. 45, 1–28.

Lindahl, E., 1919. Die Gerechtigkeit der Besteuerung. Gleerup, Lund, Translated in part as: Just taxation—A positive solution. In: Musgrave, R.A., Peacock, A.T. (Eds.), 1958. Classics in the Theory of Public Finance. Macmillan, London, pp. 168–176.

Lindahl, E., 1928. Einige strittige Fragen der Steuertheorie. In: Mayer, H. (Ed.), Die Wirtschaftstheorie der Gegenwart, vol. IV. Julius Springer, Vienna, pp. 282–304. Translated as: Some controversial questions in the theory of taxation. Musgrave, R.A. Peacock, A.T. (Eds), 1958. Classics in the Theory of Public Finance. Macmillan, London, pp. 214–232.

Lydall, H.F., 1959. The distribution of employment incomes. Econometrica 27, 110–115.

Malthus, T.R., 1798. An Essay on the Principle of Population. J. Johnson, London, Pelican Classics Edition, edited by Anthony Flew. Penguin Books, Harmondsworth, 1970. ("First Essay.").

Malthus, T.R., 1803. An Essay on the Principle of Population. J. Johnson, London, Cambridge Texts in the History of Political Thought, edited by Donald Winch, Cambridge University Press, Cambridge, 1992. ("Second Essay.").

Marshall, A., 1890. Principles of Economics, eighth ed. Macmillan, London, 1920.
Marshall, A., 1919. Industry and Trade. Macmillan, London.
Marx, Karl (1867–1894), Das Kapital. vol. 1, 1867, vol. 2, 1885, vol. 3, 1894. Hamburg, Meissner. Abridged translation as Capital. Oxford World Classics, Oxford (1995).
Meade, J.E., 1964. Efficiency, Equality and the Ownership of Property. Allen & Unwin, London.
Menger, C., 1871. Grundsätze der Volkswirtschaftslehre. Braumüller, Wien. Translated as: Principles of Economics. Free Press, Glencoe, IL, 1950.
Mill, J.S., 1848. Principles of Political Economy, Collected Works of John Stuart Mill, vols. 2–3. University of Toronto Press, Toronto, 1965.
Mill, J.S., 1863. Utilitarianism, Collected Works of John Stuart Mill, vol. 10, Essays on Ethics, Religion, and Society. University of Toronto Press, Toronto, 1969.
Mill, J.S., 1869. On the subjection of women, Collected Works of John Stuart Mill, vol. 21, Essays on Equality, Law, and Education. University of Toronto Press, Toronto, 1984.
Mirrlees, J.A., 1971. An exploration in the theory of optimum income taxation. Rev. Econ. Stud. 38, 175–208.
Morgan, M.S., 1990. The History of Econometric Ideas. Cambridge University Press, Cambridge.
Mossin, J., 1968. Taxation and risk-taking: an expected utility approach. Economica 35, 74–82.
Musgrave, R.A., 1959. The Theory of Public Finance. McGraw-Hill, New York.
Musgrave, R.A., Peacock, A.T. (Eds.), 1958. Classics in the Theory of Public Finance. Macmillan, London.
Nash Jr., J.F., 1950. The bargaining problem. Econometrica 18, 155–162.
Nozick, R., 1974. Anarchy, State, and Utopia. Basic Books, New York.
O'Brien, D.P., 2004. The Classical Economists Revisited. Princeton University Press, Princeton.
Pareto, V., 1895. La legge della domanda. Giornale degli Economisti 2 (10), 59–68.
Pareto, V., 1896a. La curva delle entrate e le osservazioni del prof. Edgeworth. G. Econ. Ann. Econ. 2 (13), 439–448.
Pareto, V., 1896-1897b. Cours d'économie politique. Rouge, Lausanne.
Pareto, V., 1909. Manuel d'économie politique. Giard & Brière, Paris, Translated as: Manual of Political Economy. Macmillan, London, 1971.
Phelps Brown, H., 1977. The Inequality of Pay. Oxford University Press, Oxford.
Phelps Brown, E.H., Browne, M.H., 1968. A Century of Pay. Macmillan, London.
Pigou, A.C., 1920. The Economics of Welfare, fourth ed. Macmillan, London, 1952.
Ranadive, K.R., 1978. Income Distribution. The Unsolved Puzzle. Oxford University Press, Bombay.
Rawls, J., 1972. A Theory of Justice. Oxford University Press, Oxford.
Redford, A., 1926. Labour Migration in England. Manchester University Press, Manchester.
Ricardo, D., 1817. On the Principles of Political Economy and Taxation, third ed. John Murray, London, 1821. The Works and Correspondence of David Ricardo, edited by Piero Sraffa, vol 1. Cambridge University Press, Cambridge, 1951.
Robbins, L., 1932. An Essay on the Nature and Significance of Economic Science. Macmillan, London.
Robinson, J., 1933. The Economics of Imperfect Competition, second ed. Macmillan, London, 1969.
Roy, A.D., 1950. The distribution of earnings and of individual output. Econ. J. 60, 489–505.
Roy, A.D., 1951. Some thoughts on the distribution of earnings. Oxf. Econ. Pap. 3, 135–146.
Sahota, G.S., 1978. Theories of personal income distribution: a survey. J. Econ. Lit. 16, 1–55.
Samuelson, P.A., 1947. Foundations of Economic Analysis. Harvard University Press, Cambridge, MA.
Samuelson, P.A., 1953. Prices of factors and goods in general equilibrium. Rev. Econ. Stud. 21, 1–20.
Sandmo, A., 1999. Asymmetric information and public economics: The Mirrlees-Vickrey Nobel Prize. J. Econ. Perspect. 13 (1), 165–180.
Sandmo, A., 2011. Economics Evolving. Princeton University Press, Princeton, NJ.
Schultz, T.W., 1961. Investment in human capital. Am. Econ. Rev. 51, 1–17.
Sen, A., 1973. On Economic Inequality. Oxford University Press, Oxford.
Senior, N.W., 1836. An Outline of the Science of Political Economy. Clowes, London.
Simon, H.A., 1957. The compensation of executives. Sociometry 20, 32–35.
Smith, A., 1759. The Theory of Moral Sentiments. A. Millar, London, Edinburgh: A. Kincaid and J. Bell. Glasgow Bicentenary Edition, edited by D. D. Raphael and A. L. Macfie. Oxford University Press, Oxford, 1976.

Smith, A., 1776. An Inquiry into the Nature and Causes of the Wealth of Nations. Strahan and Cadell, London, Glasgow Bicentenary Edition, edited by R. H. Campbell and A. S. Skinner. Oxford University Press, Oxford, 1976.

Stigler, G.J., 1941. Production and Distribution Theories: The Formative Period. Macmillan, New York.

Stiglitz, J.E., 1969. Distribution of income and wealth among individuals. Econometrica 37, 382–397.

Stolper, W.F., Samuelson, P.A., 1941. Protection and real wages. Rev. Econ. Stud. 9, 58–73.

Veblen, T., 1899. The Theory of the Leisure Class. Macmillan, New York.

Veblen, T., 1904. The Theory of Business Enterprise. Scribner, New York.

Vickrey, W., 1945. Measuring marginal utility by reactions to risk. Econometrica 13, 319–333.

von Neumann, J., Morgenstern, O., 1947. Theory of Games and Economic Behavior, second ed. Princeton University Press, Princeton.

von Thünen, J.H., 1826, 1850, vols. 1–2. Translated in part as *von Thünen's Isolated State*. Pergamon Press, Oxford, 1966.

Walras, L., 1874-1877. Éléments d'économie politique pure. L. Corbaz, Lausanne, Translated and with an introduction by William Jaffé as Elements of Pure Economics. Irwin, Homewood, IL, 1954.

Whitaker, J.K., 1988. The distribution theory of Marshall's Principles. In: Asimakopulos, A. (Ed.), Theories of Income Distribution. Kluwer Academic Publishers, Boston, pp. 105–132.

Whitaker, J.K., 1990. What happened to the second volume of the *Principles*? The thorny path to Marshall's last books. In: Whitaker, J.K. (Ed.), Centenary Essays on Alfred Marshall. Cambridge University Press, Cambridge, pp. 193–222.

Wicksell, K., 1896. Finanztheoretische Untersuchungen. Fischer, Jena, Translated in part as: A new principle of just taxation. In: Musgrave, R.A., Peacock, A.T. (Eds.), 1958. Classics in the Theory of Public Finance. Macmillan, London, pp. 72–118.

Wicksell, Knut (1901-1906), Föreläsningar i nationalekonomi. vol. 1 (1901), vol. 2 (1906). Lund: Gleerup. Translated as: Lectures on Political Economy. vol. 1, General Theory (1934), vol. 2, Money (1935). Routledge and Kegan Paul, London.

Wicksteed, P., 1894. An Essay on the Co-Ordination of the Laws of Distribution. Macmillan, London.

Zeuthen, F., 1928. Den økonomiske fordeling (The Economic Distribution). Nyt Nordisk Forlag, Copenhagen.

Zeuthen, F., 1930. Problems of Monopoly and Economic Warfare. Routledge, London.

CHAPTER 2

Inequality, Income, and Well-Being

Koen Decancq[*,†,‡], **Marc Fleurbaey**[§], **Erik Schokkaert**[†,‡]
[*]Centre for Social Policy Herman Deleeck, University of Antwerp, Antwerp, Belgium
[†]Department of Economics, Katholieke Universiteit Leuven, Leuven, Belgium
[‡]CORE, Université catholique de Louvain, Louvain-la-Neuve, Belgium
[§]Princeton University, Princeton, NJ, USA

Contents

Handbook of Income Distribution, Volume 2A
ISSN 1574-0056, http://dx.doi.org/10.1016/B978-0-444-59428-0.00003-5

Abstract

Individual well-being depends not only on income but also on other dimensions of life, such as health, the quality of social relations and of the environment, employment, and job satisfaction. In this chapter we survey the economic literature on how to construct such overall measures of well-being. We distinguish three approaches: the capability (and functionings) approach, the use of subjective life satisfaction measures, and the calculation of equivalent incomes. We discuss the normative assumptions underlying these three approaches, focusing on two issues: the degree to which individual preferences are respected and where in each approach the boundaries of individual responsibility are drawn. We compare the measurement of inequality in well-being with the use of multidimensional inequality measures. We illustrate the general theoretical issues in three domains of application: measuring the effects of household size and composition in the literature on equivalence scales, valuing publicly provided goods and services, and making international comparisons of well-being involving international purchasing power parity comparisons.

Keywords

Well-being, Capability approach, Life satisfaction, Equivalent income, Multidimensional inequality, Equivalence scales, Publicly provided services, Purchasing power parities

JEL Classification Codes

D31, D63, I30

2.1. INTRODUCTION

The economic literature on inequality has traditionally focused on income inequality. One reason to be interested in income inequality is that it may be linked to potential economic growth, to aggregate consumption, and to the occurrence and size of cyclical movements (see Chapter 14). From this perspective, income (in)equality is instrumental to reach other social objectives. A second reason is a normative one, considering the distribution to be a matter of social concern in itself, independent of its effects on other variables. In this chapter, we focus on the latter reason, leaving the instrumental concerns aside.

One normative reason to be concerned with income distribution is that we are ultimately interested in the distribution of well-being and that we consider income as a proxy for well-being. A related argument emphasizes the right of everyone to have access to a minimum level of resources, income then being an indicator of these resources. These two approaches are closely related if we define well-being directly in terms of resources, but they may differ if one adopts alternative definitions of well-being, for instance, in terms of functionings or capabilities or in terms of subjective satisfaction with life. A third normative reason why people are interested in income distribution has to do with the fairness of the process through which income is acquired. There are strong convictions in society that individuals should be paid in a fair way and that effort should be somehow rewarded. There is more discussion about the ethical desirability of remunerating productivity differences stemming from differences in innate talent or in socioeconomic background. These considerations are linked to the debate about the content of "desert" and "merit" and their relevance for evaluating the income distribution. At first sight, such a focus on the process of income formation is very different from a concern for the final distribution of well-being. Yet, although the latter is the connecting thread of this chapter, fairness judgments will also play an important role at some points in our discussion.

It is not obvious that income is indeed an adequate proxy for well-being. It is well accepted that the same monetary amount may yield a different level of well-being for individuals with different needs. Moreover, individuals do not care only about their income. A consensus seems to be emerging that information on other dimensions of life (such as health, job quality, the natural and social environment in which people are living) should be integrated into a richer view of well-being (see Stiglitz et al., 2009, and the references therein). This broadening of the perspective on well-being has led to a growing aversion against the use of (even a "corrected" or "extended") income metric to measure well-being on the ground that this would reflect a kind of "resource fetishism." Yet, from an applied viewpoint, monetary measures have the obvious advantage that they yield an operational and cardinal measure.

This brings us to the main question for this chapter: is it possible to formulate an ethically attractive notion of individual well-being that is richer than monetary income and that is still sufficiently operational to be used in applied welfare analysis?[1]

The concept of individual "well-being" can be approached from many different perspectives. One could, e.g., take a psychological perspective and investigate what is the best measure of well-being for describing and explaining the emotion of "feeling well" (Kahneman and Krueger, 2006). However, according to the welfare economic perspective taken in this chapter, the choice of an adequate measure of individual well-being is not a psychological, but rather a normative question. An adequate measure of well-being makes interpersonal comparisons such that redistribution from a better-off individual to a worse-off individual yields a better state of affairs as seen from the social welfare point of view.[2] In other words, an adequate measure of well-being serves as equalizandum for an egalitarian policy.

[1] In a macro setting, the criticism of "income fetishism" is voiced even more loudly against the use of GDP (growth) as an indicator of welfare. We do not go into the debate on sustainability or on the different ways of "correcting" GDP to include distributional issues. A critical discussion of different approaches can be found in Fleurbaey and Blanchet (2013). In this chapter we focus on the derivation of a measure of well-being at the individual level. However, in Section 2.5 we will discuss some of the implications of our discussion for international welfare comparisons.

[2] When we talk about redistribution here, we refer to redistribution of well-being. This is not necessarily restricted to income redistribution, but may also take place through spending more or less on health care or education for specific groups in the population.

The choice of a particular metric of well-being is inevitably a matter of value judgments. Selecting income as the measure of well-being, or deciding to go beyond the income dimension is a normative choice. Indeed, the argument that we should include other dimensions than income because people care about these other dimensions, for instance, is only valid if we accept the normative position that society should care about what people care about.

Because defining an adequate concept of well-being is a normative choice, it is no surprise that opinions differ about what is the best measure of well-being. As soon as one moves beyond the single income dimension to describe well-being, at least two sets of issues come to the fore. First, what additional dimensions should be included? Or, more fundamentally, what justification can or should be given for this choice of dimensions? Second, should these different dimensions be seen as incommensurable, or is it possible to aggregate them into one measure of individual well-being? If one takes the former position and sticks to a vector representation of well-being, how should one handle interpersonal comparisons involving a trade-off between the different variables? If one takes the latter position, what should be the normative logic underlying the aggregation across dimensions?

In this chapter, we will describe the answers proposed by different approaches to these two sets of questions, and we will discuss their normative implications. In particular, we will look at the different approaches from two specific perspectives. First, we focus on the extent to which the proposed measure of individual well-being respects individual preferences. The principle of individual sovereignty has always been one of the main tenets of economics and remains a hotly debated issue (see, among others, Hausman and McPherson, 2009). One of the difficulties in the debate is that different interpretations have been given to the concept of preferences. In this chapter, we will interpret preferences as reflecting people's well-informed and well-considered ideas about what is a good life. The recent literature has documented many behavioral anomalies and has convincingly shown that these well-informed preferences are not always revealed in actual choice behavior (see Della Vigna, 2009, for an overview). Preferences as the representation of a life project therefore do not coincide with the traditional economic concept of revealed preferences. We will come back to these distinctions later.

Second, we will focus on the implications of selecting a specific measure of well-being for the delineation of the domain of personal responsibility. If we start from the idea that a redistribution from someone with a higher level of well-being to someone with a lower level of well-being is an improvement from the social point of view, this implies (even when this is not made explicit) that the person at the lower level of well-being is not held responsible for this lower level of well-being. This responsibility perspective helps to interpret some of the normative differences between the various approaches. As responsibility and freedom are closely related concepts, a focus on responsibility will also allow us to comment on the different interpretations of freedom embedded in the different approaches.

The chapter is structured as follows. Section 2.2 gives a brief historical sketch of the development of the literature. Section 2.3 is the core of the chapter. It contains a critical discussion of the three prominent proposals for a measure of well-being capturing nonincome dimensions: the capabilities approach, the subjective well-being approach, and the equivalent income approach. In Section 2.4 we consider the literature on multidimensional inequality measurement and on multidimensional stochastic dominance, which has taken up a direct concern for the distribution in multiple dimensions without introducing explicitly a measure of individual well-being. We refer to Chapter 3 in this volume for more details on the different indices that have been proposed, and we focus on their theoretical foundations in the light of our normative criteria. In Section 2.5, we apply the general insights from Sections 2.3 and 2.4 to a series of issues that have played a prominent role in the applied literature: the use of equivalence scales to deal with heterogeneous households, the inclusion of the value of public and nonmarketed goods and services in the measurement of inequality, and the measurement of inequalities at the world level. This last point will also give us the opportunity to link the discussion on the limitations of gross domestic product (GDP) as a measure of aggregate social welfare to the normative issues discussed in the previous sections.

Before starting, we make two remarks. First, this chapter is about evaluating states of affairs. For such a broad evaluation, the income distribution is not sufficient; it is necessary to work with a broader concept of well-being. This does not mean that it would not be relevant to argue in favor of a *re*distribution of income. In fact, as soon as we define well-being to include personalized and nontransferable characteristics (such as health), a direct redistribution of well-being is not feasible. If income has a positive effect on well-being, a redistribution of income can be an effective instrument to realize a more equal distribution of well-being. Yet, this will not necessarily be a redistribution from the (income-)rich to the (income-)poor because the income-rich can be at a lower level of well-being than the income-poor. This is precisely where different approaches to well-being will make a difference.

Second, we will focus on inequality rather than on poverty. The two concepts are complementary, but may still involve different ethical intuitions, especially if one accepts that poverty has an absolute component. It is probably not a coincidence that the concept of multidimensional deprivation has traditionally played a more important role in poverty than in inequality research. Poverty researchers are generally more sympathetic toward rights-based approaches. They are also less inclined to accept the idea that different life dimensions can be traded off. We will discuss these issues where needed.

2.2. A BRIEF HISTORICAL SKETCH

As we defined the quest for a measure of well-being ultimately as a quest for an attractive equalizandum for egalitarian policy, our discussion is related to the large welfare economic literature on consequentialist versus nonconsequentialist approaches and within the consequentialist approach on welfarist versus nonwelfarist criteria. Although this literature is rich and inspiring, it is also plagued by some terminological confusion. This problem is, e.g., very acute for the (for our purposes essential) notion of "welfarism." We will therefore briefly situate our topic against the background of the debate around Arrow's (1951) impossibility theorem. We will do so in an informal way without going into technical details.[3]

As a matter of fact, the initially dominant interpretation of Arrow's theorem was that it was impossible to define a nondictatorial social ordering of social states, satisfying the Pareto principle of respect for individual preferences. It was soon realized, however, that the independence condition, which was necessary to arrive at the impossibility result, was a strong one. It basically stated that the social ranking of any pair of two alternatives should depend only on the ordinal noncomparable individual preferences over these two alternatives. The so-called informational approach to social choice (d'Aspremont and Gevers, 1977; Sen, 1970)[4] showed that the impossibility is lifted as soon as one accepts that it is meaningful to represent individual preferences with an interpersonally comparable utility function. Depending on the specific informational assumptions made, a whole range of social orderings can then be defined, ranging from the utilitarian sum of utilities to the leximin ordering giving priority to the worst off. The lesson seemed to be that the only way to escape from Arrow's impossibility was to work with such an interpersonally comparable notion of utility (i.e., to go beyond ordinal noncomparable individual preferences).

Amartya Sen, who was one of the main contributors to this literature, later became one of the main critics of what he coined as welfarism (i.e., the approach in which the social evaluation is based solely on individual but interpersonally comparable levels of subjective well-being; see, e.g., Sen and Williams, 1982). Important social philosophers had already rejected the welfarist approach. Rawls (1971, 1982) stated that individuals have life projects and that these life projects should be respected, but that it does not make sense to reduce them to the objective of reaching the maximum level of welfare. What matters to individuals is the content of their project, not the satisfaction following from its realization. These projects are incommensurable. Dworkin (1981a) emphasized the problem of expensive tastes; according to him, someone with expensive tastes (e.g., for "prephylloxera claret and plover's eggs," as in Arrow's (1973) famous example)

[3] Pattanaik and Xu (2012) offered a conceptual framework to structure the various approaches. See also Fleurbaey (2003) for a more formal treatment.

[4] See d'Aspremont and Gevers (2002) for a survey.

cannot claim that he should be compensated for his ambitions at the expense of those with more modest tastes. Sen (1985) reformulated similar arguments in an elegant way by pointing out that subjective welfarism suffered from two problems. The first he calls "physical-condition neglect": Utility is only grounded on the mental attitude of the person and does not sufficiently take into account the real physical conditions of the person. This has two aspects. One is the issue of expensive tastes; the other is that persons may adapt to their objective circumstances or realistic expectations: "A person who is ill-fed, undernourished, unsheltered and ill can still be high up in the scale of happiness or desire-fulfillment if he or she has learned to have 'realistic' desires and to take pleasure in small mercies" (Sen, 1985, p. 21). The second problem is "valuation neglect." Valuing a life is a reflective activity in a way that "being happy" or "desiring" need not be (Sen, 1985, p. 29). An acceptable approach to well-being should explicitly take into account this valuational activity by the persons themselves.

In a long series of books and papers, Sen proposed his own concept of well-being in terms of functionings and of "advantage" in terms of capabilities. This approach is the first notion of well-being that we discuss in more detail in Section 2.3. It is definitely non-welfarist in Sen's own original meaning of the word, as it does not interpret well-being in terms of subjective welfare. It is even explicitly formulated as an alternative to subjective welfare. Yet, it does evaluate social states in terms of the individual achievements (the individual advantage levels) in these social states. It is therefore consequentialist and individualist. And here the terminological confusion starts. Some authors (e.g., Pattanaik and Xu, 2012) claim that it is natural to use the term *welfarist* also for such approaches that are centered on personal well-being, even if they use a well-being concept that is *not* subjective utility. Although we prefer the original (narrower) use of the term by Sen, we will try to avoid confusion by adding the word *subjective* each time we use the notion of welfarism in this original meaning.

At a time when the criticism on subjective welfarism was winning ground among welfare economists, there was a surprising and spectacular growth of the interest in the measurement of happiness in other domains of economics. Advances in survey research suggested more and more convincingly that happiness and/or life satisfaction could be measured and that interpersonal comparisons of these measured concepts yielded meaningful results. A rapidly growing stream of empirical papers showed that life satisfaction is not exclusively determined by income but is also strongly influenced by nonmonetary dimensions of life (such as health, social interactions, and job market status). The econometric results were reasonably robust. Although a large part of this literature is meant to be only explanatory, the implicit suggestion that what contributes to happiness must per se be good is very strong. Moreover, some authors (Frey and Stutzer, 2002; Kahneman et al., 1997, 2004; Layard, 2005) have been explicit about the normative implications of their empirical work: Now that we know how to measure utility, why not go back to Bentham? This position reflects a remarkable revival of subjective welfarism, and it is striking that the happiness literature has largely disregarded the arguments against subjective welfarism of the philosophical and welfare economic literature. The life satisfaction approach is the second notion of well-being that will be discussed in Section 2.3.

The third notion of well-being that we will explore in detail is that of "equivalent income," or money-metric utility. It also has a somewhat surprising history. Money-metric utility was introduced as a representation of preferences by Samuelson (1974) and Samuelson and Swamy (1974) and had some impact on the applied welfare economic literature during the eighties (see Deaton and Muellbauer, 1980; King, 1983, for instance). It lost popularity, however, as authors argued that it relied on an arbitrary choice of reference values and could have nonegalitarian implications (Blackorby and Donaldson, 1988). Although it slowly disappeared from the applied welfare economic literature, it was (more or less independently) developed within the social choice literature in what is called the theory of fair allocation. This theory looked for a social ordering that was based only on noncomparable ordinal preferences (i.e., noncomparable life projects). At first sight, this attempt may look hopeless because it should run against Arrow's impossibility result. However, closer investigation shows that Arrow's independence axiom can be decomposed in two components (Fleurbaey and Mongin, 2005; Roemer, 1996): The first is "ordinal noncomparability," stating that the only information that can be used is information about individual ordinal preferences; the second is "binary independence," requiring that the ranking of two alternatives should depend only on the individual evaluation of these two alternatives (see Fleurbaey and Blanchet, 2013, p. 139). The welfarist approach relaxes the first component, the fair allocation approach the second. Going beyond binary independence makes it possible to use information about the indifference curves for the two alternatives. Moreover, it turns out that the theory of fair allocation has a (more or less convincing) reply to the criticism that had been raised against the use of money-metric utilities. We summarize this debate in Section 2.3.

Note that the equivalent income is yet another concept of individual well-being that does not coincide with subjective welfare, but is based on individual preferences. Again, some authors claim that "respecting individual preferences" boils down to welfarism. This is then a third possible interpretation of the term (the first relating to the use of comparable subjective utility levels, the second to any measure of personal well-being). As mentioned already, we will not follow this line of thinking, and we will reserve the use of the term *subjective (non)welfarism* to its original meaning.

Although the three approaches that we sketched until now can be related more or less explicitly to the welfare economic literature on the measurement of well-being, this is much less the case for another strand of economic research that aimed at broadening the concern for income inequality to include other dimensions. This literature deliberately remains agnostic about the formulation of an individual measure of well-being and focuses directly on Pigou–Dalton axioms in a multidimensional space. We will discuss the relationship between this approach and the welfare economic literature in Section 2.4.

2.3. INEQUALITY OF WHAT?

We first discuss the approach that defines well-being in terms of functionings and capabilities, then consider the normative interpretation of happiness and life satisfaction data and, finally, turn to the equivalent income approach. In each case we investigate whether the well-being concept respects individual preferences and what the underlying (implicit or explicit) delineation of individual responsibility is.

It is useful to introduce some notation. Let ℓ_i denote the vector of m aspects of life that may matter to individual i. Examples are consumption or income, health, longevity, leisure, status, and job characteristics. One of the variables in ℓ_i is income (or consumption) y_i. Individuals have a life project (i.e., an informed judgment about what makes a life good or bad). We represent this life project for each individual i by a preference ordering R_i over the vectors ℓ_i: $\ell_i R_i \ell_i'$ if i weakly prefers the life described by ℓ_i to the life described by ℓ_i'. Let $\ell_i P_i \ell_i'$ denote strict preference and $\ell_i I_i \ell_i'$ denote indifference. These well-informed preferences are individual specific. We do not assume that these preferences are always revealed in actual choices. Subjective individual satisfaction is given by a "satisfaction function" $S_i\,(\ell_i)$.[5]

We assume that, from a normative point of view, individual i's situation is completely described by the triplet (ℓ_i, R_i, S_i), that consists of the vector of life dimensions ℓ_i, the preference ordering R_i and the satisfaction function S_i. This means that personal characteristics (e.g., cognitive capacities) are only relevant insofar as they influence preferences or satisfaction, or if they are part of the vector of relevant life dimensions for individual i. A method of interpersonal well-being comparisons must be able to rank such triplets (ℓ_i, R_i, S_i). Because ℓ_i by definition describes all the aspects of life that matter to individual i, measuring i's well-being involves constructing an interpersonally comparable index in which the various elements of ℓ_i are weighted. Different well-being concepts are represented by a well-being measure WB. The value $WB(\ell_i, R_i, S_i)$ is to be interpreted as the well-being of individual i with life ℓ_i, preference ordering R_i and satisfaction function S_i.

2.3.1 Functionings and Capabilities

The origins of the capability approach within welfare economics are to be found in a series of influential papers and monographs, written by Amartya Sen in the 1980s (Sen, 1980, 1985; Sen et al., 1987). He developed and discussed the approach further

[5] We do not use the term *utility function* here because this refers to *any* representation of the ordinal preference ordering. The "satisfaction function" is one choice out of all the possible utility functions. As we will see, the equivalent income function is another.

in some widely read books (Nussbaum and Sen, 1993; Sen, 1992, 1999, 2009). On the philosophical side, important contributions have been made by Nussbaum (2000, 2006, 2011). Both authors, and the many papers following in their wake, are explicit about the normative purpose of the approach. Their aim is to define individual well-being so that it can be used in a meaningful way as the equalizandum for an egalitarian policy. The important question is: "Equality of what?" (Sen, 1980).

Sen's own answer to this question starts from the rejection of two extreme alternative approaches. We have seen in the previous section that Sen considers subjective welfarism unacceptable because of the problems of "physical condition-neglect" and "valuation neglect." But focusing exclusively on either income or material resources would not do justice to the heterogeneity among human beings, he argued. What does matter to define well-being is the vector of functionings of a person (i.e., achievements): what this person manages to do or to be (such as being well nourished, well clothed, mobile, or able to appear in public without shame).

These functionings have to be distinguished from the resources or commodities that are used to achieve them. Personal and environmental characteristics, to a large extent, determine what people can achieve with a given amount of resources. How well nourished a person is does not only depend on the amount of food eaten, but also on the biological characteristics of the person's body and the work that person does; books do not contribute to the personal development of persons who were never taught to read; whether a person is mobile does not only depend on whether that person owns a bicycle, but also on the availability of safe road infrastructure, and so on. The well-being of person i can be seen as the person-specific valuation of the vector of functionings ℓ_i:

$$\mathrm{WB}^{\mathrm{F}}(\ell_i, R_i, S_i) = v_i(\ell_i), \tag{2.1}$$

where the superscript F refers to the functionings approach. The crucial question is, of course, how to interpret the valuation function v_i. We return to this question later in this section.

In a further step, Sen claimed that a description of well-being in terms of achieved functionings is not sufficient because it does not integrate the essential notion of freedom. His classical example involves the comparison between two individuals who are both undernourished. For the first person, the undernourishment is the result of material deprivation. The second person, on the contrary, is wealthy but freely chooses to fast for religious reasons. Although their achievements in terms of the "being well-nourished" functioning are exactly identical, it is intuitive to say that their situations are not the same in terms of well-being. Therefore Sen introduced the notion of "capabilities" to capture the real opportunities of persons. The capabilities of person i are given by the set of functionings vectors that are accessible to the person (i.e., the set from which he or she can choose). Loosely formalized, we can represent the set as $Q_i = \{\ell_i | \ell_i \text{ is feasible for individual } i\}$. The "advantage" of person i is then the evaluation of the capability set Q_i.

We will now first discuss the implications of moving from functionings to capabilities (i.e., from achievements to opportunities).[6] This will allow us to discuss the interpretation of freedom and responsibility within the capability approach. We then move to the issues of choosing the relevant dimensions to be included in the vector ℓ_i, and whether and how to aggregate them in a single well-being indicator. This will clarify the position of the capability approach with respect to the other normative criterion (i.e., respecting individual preferences).

2.3.1.1 Capabilities, Responsibility, and Freedom

Capabilities, defined as the opportunity set from which people can choose, are a reflection of the real (positive) freedom of individuals and are definitely not restricted to the securing of negative freedoms alone. People should not only have the legal right to provide themselves with food, but they should also have the economic possibilities to do so. Equalizing capabilities also goes beyond eliminating discrimination, although the latter is an important element of it. This integration of positive freedom issues in the measurement of well-being is an attractive idea. However, it also raises some difficult questions.

A first issue was raised by Basu (1987) in his review of Sen (1985) and was taken up again in Basu and Lopez-Calva (2011). It can best be illustrated in the usual Edgeworth box of a two-person two-good exchange economy (see Figure 2.1). This figure depicts a general equilibrium situation (point e), in which relative prices are given by the slope of the line AB, and the initial endowments of persons 1 and 2 are given by point a. In this setting it might seem straightforward to say that individuals choose within their budget sets (i.e., the areas O_1CAB for person 1 and O_2DBA for person 2). But the figure immediately shows that their freedom to choose within the budget set is illusory: What is open to one person depends on what the other person chooses. If person 2 sticks to the bundle in e, it is impossible for person 1 to pick bundle b. In fact, in that case person 1 can only choose bundles from the rectangle O_1FeG. In general terms, changes in the choices by one person (induced by changes in preferences, for instance) will change relative prices and, therefore, the opportunity set of the other person. Figure 2.1 represents the very peculiar case of a two-person two-good exchange economy, but the point made by Basu is more general. The achieved functionings of any person do not only depend on the choices made by that individual, but also on actions taken by other individuals. How to define the capability set of any person in such a situation?[7]

[6] There is some terminological confusion in the literature. Although the distinction between "functionings" and "capabilities" is clear in Sen's approach, other authors have later used the term (*basic*) *capabilities* to refer to functionings. In this chapter we aim to remain as close as possible to the original meaning of the terms.

[7] Basu's example of the Edgeworth box loses some of its relevance in a many-person society, but the issue of social interdependencies is a more general one. See Pattanaik and Xu (2009) for a discussion leading to the conclusion that none of the solutions proposed for this problem is entirely satisfactory.

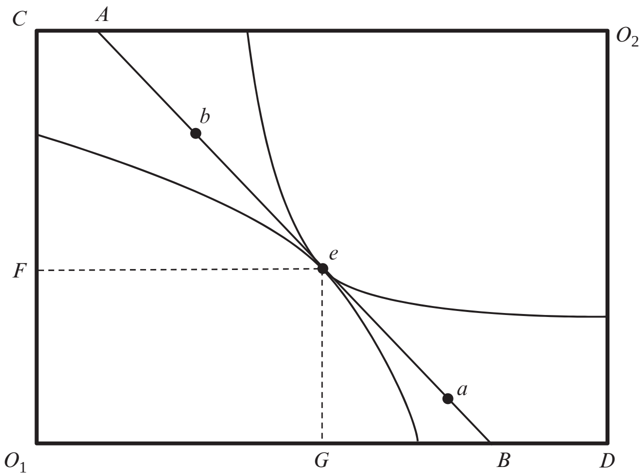


Figure 2.1 Capabilities and social interactions.

In general, defining well-being in terms of opportunity sets requires that one can put a value on these sets in a normatively attractive way. This is a difficult problem, as demonstrated by the formal (and abstract) literature on the topic (see, among others, Barbera et al., 1998; Foster, 2010). If one does not include information about preferences, a set of reasonable axioms soon leads to the unattractive solution of evaluating opportunity sets by simply counting the number of its elements without taking any account of the "quality" of these elements (Pattanaik and Xu, 1990). Yet, introducing preferences does not lead to easy solutions either. One of the proposals by Sen (1985) is to evaluate sets by the value of their best element. He called this the "elementary evaluation," but immediately acknowledged that this method does not do justice to the idea of freedom. Another proposal would be to say that the set Q is "better" than Q' if there is an element in Q that is considered by all individuals to be better than all elements in Q'. This is a very strict criterion that leaves many sets incomparable when there is sufficient interpersonal heterogeneity in preferences. At this moment, there seems to be no single proposal that has gained enough theoretical support to become the prime candidate to be implemented in applied work. As a matter of fact, except for the simple proposals such as elementary evaluation, none of the proposed methods appears easy to operationalize. This lack of practical applicability is worrying if we opt for "advantage" as our preferred measure of well-being.

An even more fundamental problem with the proposals to value sets that have been made in the literature until now is that the proposed methods go against a compensation principle that has been a cornerstone of the theory of equality of opportunity (Fleurbaey and Blanchet, 2013, p. 218). This principle says that individuals that are at the same level of individual responsibility (say, effort) should also obtain an equally valuable outcome. The literature on equality of opportunity is surveyed in Chapter 4.

Another important question is the following. If we want to measure well-being, is it then sufficient to look at opportunity or capability sets while neglecting completely the realized achievement or functioning? Fleurbaey (2006a) argued that focusing exclusively on capability sets suffers from two problems. First, it leads to a loss of information. Consider the following example. Two individuals, Ann and Bob, face the capability sets Q and Q', respectively, with $Q \subset Q'$. Bob has all the opportunities that Ann has, and he even has some opportunities that are not feasible for Ann. In such a case of set inclusion, it seems reasonable to say that Bob's opportunities are at least as good as Ann's. But now suppose that Ann selects option a in Q, while Bob picks option b in Q', such that $a \gg b$. Then it seems reasonable to say that Ann's achievement is "better" than Bob's. Such an example is not irrelevant within the capability approach, as many of its advocates stress that individuals do not necessarily choose within their capability set the functioning vector that would give them the highest level of individual well-being. Whatever conclusion one wants to draw from this example in terms of who is worse off, it is clear that limiting our attention to sets and neglecting actual achievements leads to a loss of information. Comparing sets and knowing the selected option is not the same as comparing sets without knowing the selected option.

Second, an evaluation purely on the basis of capability sets may reflect a harsh attitude about individual responsibility, given the well-documented limitations of individual decision-making capacities. With capabilities as the measure of well-being, individuals are held responsible for their mistakes when selecting a particular option from their capability set. The question gets even more pressing when considering choices over the life cycle. The opportunity sets of older people are determined by their decisions when they were young, and the question arises for how long individuals have to remain responsible for potential "mistakes" committed earlier in life.

The previous discussion raises the question of whether focusing on capability sets is indeed the best way to introduce freedom considerations into the measurement of well-being. From his first writings on the topic, Sen has hinted that an alternative is to work with so-called refined functionings or comprehensive outcomes, where the "refinement" refers to the operation of including additional information on the available alternatives or on the process of choice itself. Let us reconsider the example of the two individuals who are fasting and starving. The fasting person is choosing to eat less; the poor starving person is exercising no choice at all. These can be seen as two different "refined" functionings—choosing a when b is also available is a different refined functioning than choosing a when b is not available (Sen et al., 1987, pp. 36–37). Alternatively, in addition to the functioning of being well nourished, one could consider another functioning "exercising choice with respect to what one eats."

Fleurbaey (2009) extends this idea and argues that all the relevant aspects of freedom can be captured through refined functionings. Basic freedoms of thought, speech, political activity, travel, and so on are clearly part of the functioning vector. The freedom from avoidable disease can be approximated in terms of the achieved health functioning, of the accessibility of the health care system, and of the environmental and social factors influenced by public health policy. These examples immediately show that the refined functionings approach too raises formidable challenges. Understanding the "process of choosing" is not straightforward. As soon as one has to resort to indirect indicators (such as education, income, social relations, accessibility of the health care system), it is important to carefully consider the specific social, environmental, and individual variables that determine the influence of these indicators. In moving from "capability sets" to "refined functionings," we replace the problem of evaluating sets with the challenge of understanding the process of "producing" refined functionings. However, it seems that the notion of refined functionings is better suited for a careful empirical analysis, which is needed to answer these questions about choice, well-being, and differences in opportunities.

2.3.1.2 Choice of Dimensions

Whether we prefer a definition of well-being based on capabilities or (refined) functionings, inevitably we face the question of how to select the list of relevant dimensions. It follows from a focus on freedom and agency that only dimensions that people have reason to value should be included. Yet, this notion of "reason to value" can be interpreted in different ways.

A natural choice in a freedom perspective is to include all dimensions that are considered by the individuals themselves to be relevant within their own personal life projects. This matches the normative purpose of respecting individual preferences. From the applied point of view, however, it raises the question how one should collect the necessary information about these preferences. There is a conceptual distinction between very specific dimensions (such as "not suffering from malaria") and more encompassing dimensions (such as "having a good overall health situation"). Although a consensus could perhaps be reached on the relevant dimensions at a more encompassing level, substantial disagreement may remain about the dimensions to be included at a more specific level. Direct surveys should therefore be structured carefully.[8]

Most researchers within the capability approach do not follow this preference-based approach. Their suspicion toward preferences can at least partly be explained by the

[8] Clark (2005) investigated through a small number of high-quality interviews how the South African poor perceive "development" (a good form of life). He concluded that the intrinsic value of material things matters a lot. A challenging example is Coca-Cola, which turns out to be very important to many poor respondents. Although the nutritional value of Coca-Cola is low, it is "perceived as a superior first world product" (Clark, 2005, p. 1353) and is important "to achieve other important functionings such as relaxing, facilitating social life and enhancing friendships" (Clark, 2005, p. 1354). But is "having the opportunity to drink Coca-Cola" really a crucial dimension of life?

multiplicity of preference concepts that coexist in the literature. If preferences are interpreted as revealed through actual choice behavior, a dose of suspicion is indeed justified based on the well-documented behavioral anomalies that individual choices display in real life. However, in an approach that puts great emphasis on freedom and agency, it seems less easy to discard preferences understood as reflecting the individual's well-considered life projects. One possible justification is that one considers it unrealistic to assume that individuals have such well-defined preferences. We will come back to this viewpoint later in the chapter.

Let us now describe the two alternative ways of selecting the list of relevant dimensions that have been proposed by Nussbaum and Sen, respectively. Inspired by Aristotle, Nussbaum (2000, 2006, 2011) started from an "objective" view about what constitutes human flourishing and defined a list of abstract essential capabilities (or functionings). She presented the list as universal, but is well aware of the fact that the translation of the abstract capabilities in implementable terms will depend on the specific social, cultural, and economic context. Sen, on the other hand, prefers to leave the definition of the list of functionings deliberately open, as he believes that the list should be drawn up in a democratic process through public reasoning (see, e.g., Sen, 2004). This dynamic process creates room for participation of the people concerned—which in itself is already a crucial functioning. Sen's focus on public reasoning is inspired by an activist perspective that aims at implementing the capability approach by means of social change. From an analytical and ethical point of view, however, it seems to raise many questions. If one reaches an agreement through a public deliberation process, is this agreement a kind of compromise between the different preferences of the individuals involved? If so, how is the compromise to be interpreted? Does the process not induce the risk of a tyranny of the majority or of the most outspoken personalities? If we were to accept that preferences are not given ex ante, but are formed in and through the deliberation process itself, one could perhaps even aim at a real consensus rather than a compromise. Yet, without a good understanding of (and arguably well-defined conditions imposed on) these public deliberations, it is not clear what the normative status of such a consensus should be.

Although the conceptual differences between these different approaches are important, the problem seems less acute when it comes to practical applications. Alkire (2002) gives an extensive overview of different lists of dimensions that have been proposed in the literature and reaches the following—perhaps surprising—conclusion. Despite the large variety of approaches and the differences in opinion about the underlying logic, the specific proposals are strikingly similar. As a matter of fact, the same is true for the lists of dimensions that have been proposed for practical applications by, e.g., the Organisation for Economic Co-operation and Development (OECD, 2011), the European Statistical System (2011), or Stiglitz et al. (2009). All proposals include material consumption and housing quality, health, job market status and leisure, the quality of social interactions, and the quality of the natural environment. To be precise, this consensus is about the first layer of encompassing dimensions and dissipates when we turn to a second layer of more specific dimensions. However, even at that lower level the similarities are sufficiently reassuring, if one accepts the ultimate aim of arriving at a single synthetic indicator of well-being. Indeed, (partial) overlap can be taken care of through the choice of the weights used to get at the synthetic indicator (see Decancq and Lugo, 2013 for an overview on setting weights in synthetic well-being indicators). Let us now turn to this aggregation step.

2.3.1.3 Aggregation and Respect for Preferences

Note first that the construction of a synthetic indicator of well-being is not really necessary, if the purpose of the analysis is to construct a richer description of individual well-being than is possible with a one-dimensional approach in terms of monetary income alone. In fact, for this purpose, a simple observation of the vectors ℓ_i is sufficient, and any aggregation procedure may be interpreted as leading to a loss of information. Yet, as soon as one wants to make interpersonal comparisons of well-being between all individuals of society (for instance, when computing inequality) it is necessary to go beyond the simple description in terms of vectors. In this section, we therefore focus on the construction of synthetic well-being indicators. In Section 2.4 of this chapter, we will consider approaches that introduce a multidimensional version of the Pigou–Dalton transfer principle directly at the level of the vectors of relevant life dimensions.

An influential stream within the capability approach emphatically rejects the idea that the different life dimensions are commensurable. Again, Nussbaum (2000, 2006, 2011) is the main proponent of this view. There is an immediate normative reason for this position. Nussbaum focuses on capabilities as basic needs, and she accepts the "union" identification strategy to the measurement of multidimensional poverty, in which someone is considered poor as soon as he or she does not reach a minimum level for at least one dimension. A union approach to identify the multidimensional poor is closely related to a "rights-based" view on poverty measurement. One can interpret this approach as implying a very simple ranking of individuals in which only two groups are distinguished, the poor and the nonpoor, and no further comparisons are made within these groups. This approach may be sufficient for some purposes (such as identifying the poor) but is too coarse if we want to derive conclusions about inequality in society, for instance.

If we want to derive a measure of individual well-being that can be used for the measurement of inequality, the possible trade-offs between the different dimensions can no longer be neglected. This brings us back to the interpretation of the valuation function v_i in expression (1). If individuals indeed have a continuous preference ordering over life dimensions and if one accepts the normative relevance of this preference ordering, then v_i could be a representation of their preference ordering:

$$\ell_i R_i \ell_i' \Leftrightarrow v_i(\ell_i) \geq v_i\left(\ell_i'\right).$$

Note, however, that different individuals may have different valuation functions (each representing their own personal preference ordering about what is a good life) and that, moreover, for each preference ordering there is an infinity of valuation functions that represent it (indeed, any monotonic increasing transformation of v_i is also a representation of R_i). This raises a fundamental question of interpersonal comparability. We return extensively to that problem in Section 2.3.3.

On the other hand, in the capability approach (with its suspicion for individual preferences), researchers typically aim to use a common valuation function v, which is the same for all individuals. If we do not rely on personal preferences, the question becomes how to construct such a function. Here again, we can rely on public deliberation, but this raises issues similar to those encountered when discussing the choice of relevant dimensions. Alternative, more analytical, proposals have been discussed in Sen (1985). The most prominent of these proposals is the so-called intersection approach, which makes use of a dominance principle. We can write this principle more formally, using the notation that was introduced at the beginning of this section[9]:

Dominance Principle: (ℓ', R', S') *is at least as good as* (ℓ'', R'', S'') *if* $\ell' R \ell''$ *for all* R, *and strictly better if* $\ell' P \ell''$ *for all* R.

This principle states that if a situation ℓ' is preferred to a situation ℓ'' for *all* admissible individual preference orderings (and hence also by all individuals in society), then we consider the individual in ℓ' to be better off (from a normative perspective) than the individual in ℓ'', *irrespective* of the differences in the actual preference orderings or the satisfaction functions of the individuals.[10] With monotonic preference orderings the dominance principle implies that (ℓ', R', S') is better than (ℓ'', R'', S'') whenever $\ell' \gg \ell''$.

One immediate problem with this approach (acknowledged explicitly by Sen) is that the resulting partial ordering may be very coarse. Not many triplets can be effectively ranked with respect to well-being. The deeper question, however, is why it is so difficult to obtain a more complete ordering. One answer is to say that well-being and advantage are objective concepts, and that the incompleteness follows from the fact that it is intrinsically difficult to define what a good life is. We mentioned already that this perfectionist idea is prominent in the work of philosophers in the Aristotelian tradition (most notably Martha Nussbaum). An alternative answer would be that the valuation of functionings bundles should be at least partly based on the valuations by the persons themselves (which seems to be more in line with the idea of freedom) and that the difficulty of defining a common valuation function v reflects the fact that it is not straightforward to find a kind of "overlapping consensus" on what is a good life (see, e.g., Sugden, 1993).

This difficulty is indeed fundamental. At first sight, the dominance principle appears to be in line with the respect for personal preferences. However, this first impression is misleading, as has been shown by Brun and Tungodden (2004), Fleurbaey (2007), and Pattanaik and Xu (2007) (see Weymark, 2013, for a survey). The underlying intuition is

[9] Assuming anonymity, we drop the individual subscripts to simplify notation.

[10] This latter conclusion is natural because the capability approach does not take into account subjective satisfaction for the ranking of well-being levels if it is not part of the vector ℓ.

that the dominance principle implies that (ℓ, R, S) is at least as good as (ℓ, R', S') for all ℓ and all R, R', S, S', so that preferences can play *no* role in the evaluation of (ℓ, R, S). We further illustrate the difficulty by making use of the following principle:

Personal-Preference Principle: (ℓ, R, S) *is at least as good as* (ℓ', R, S) if $\ell R \ell'$ *and strictly better if* $\ell P \ell'$.

The personal-preference principle requires that the (normative) evaluation of well-being in two situations should follow the preferences of the individual involved. As this principle only involves intrapersonal comparisons, it is a weak requirement of respect for preferences, but even this weak requirement is already incompatible with the weak form of the dominance principle stating that (ℓ, R, S) is strictly better than (ℓ', R', S') whenever $\ell \gg \ell'$. This incompatibility is shown by the following example. Figure 2.2 illustrates. Take $\ell_i, \ell_j, \ell_i', \ell_j'$ and R_i, R_j such that $\ell_i \gg \ell_j, \ell_j' \gg \ell_i', \ell_i' P_i \ell_i$, and $\ell_j P_j \ell_j'$ The personal-preference principle implies that (ℓ_i', R_i, S_i) is strictly better than (ℓ_i, R_i, S_i) and that (ℓ_j, R_j, S_j) is strictly better than (ℓ_j', R_j, S_j), whereas the dominance principle implies that (ℓ_i, R_i, S_i) is strictly better than (ℓ_j, R_j, S_j) and that (ℓ_j', R_j, S_j) is strictly better than (ℓ_i', R_i, S_i). By transitivity, this is impossible.

This incompatibility confronts us with a deep clash between two different normative principles. If one constructs a partial well-being ordering based on the idea of dominance (or consensus), one almost immediately gets a conflict with even a minimal form of respect for individual preferences. Later in this chapter we will show that it is possible to operationalize the concept of a valuation function that respects preferences and therefore necessarily violates the dominance principle. If one endorses a more objective view of well-being, this may be seen as a bridge too far.

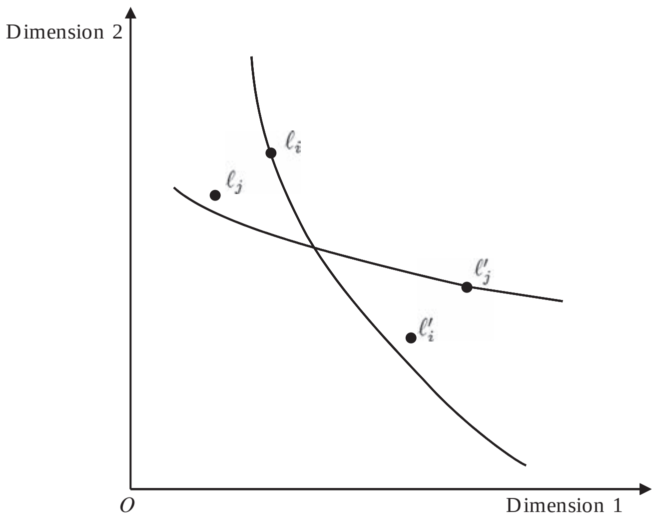


Figure 2.2 The dominance principle and personal-preference principle are incompatible.

Moving from the basic theoretical discussion to the applied work, a large number of empirical applications are focusing only on a description of functionings vectors. At the other extreme, we can find some examples in which one synthetic well-being index is constructed in an explicit way. The best known example of a synthetic well-being index (at the country level) is the Human Development Index (HDI) of the UNDP that will be described in more detail in Section 2.5. As a matter of fact, we will argue that this popular measure presents a good illustration of the problems raised with an objective approach.

2.3.2 Utility and Happiness

In the recent decades there has been a strong upsurge of economic research on happiness. Given the traditional reluctance of economists about the use of subjective information obtained through questionnaires, this is a somewhat surprising phenomenon. Data on subjective well-being have by now been collected for thousands of respondents throughout the world with large-scale surveys. Many variants of the subjective well-being question have been studied, all being relatively simple.[11]

It is useful to distinguish between two broad categories of subjective well-being questions. We illustrate both categories using the European Social Survey. The first question refers to life satisfaction and goes as follows: "All things considered, how satisfied are you with your life as a whole nowadays? Please answer using this card, where 0 means extremely dissatisfied and 10 means extremely satisfied." The second question refers to happiness: "Taken all together, how happy would you say you are? Please use this card." Analysis suggests that the results for both questions are generally highly correlated and that they can be explained to a large extent by the same set of correlates. Not surprisingly, many economists have concluded that both questions measure the same underlying concept—which is then equalized with the traditional notion of utility. This interpretation is not supported by psychologists, however, who make a clear distinction between affective and cognitive components in the experience of life satisfaction. As a matter of fact, there is growing evidence that affective scores are less sensitive to objective conditions of life, such as income, and more prone to adaptation.[12] We will return to this distinction and to its normative consequences in the first subsection, but for the moment we do as if both questions reflected the same underlying concept of utility that can either be measured by questions on happiness or by questions on life satisfaction and that we will label generically as "subjective well-being."

It is not our ambition here to give an extensive overview of the large empirical literature on the topic (see Chapter 14). We will only draw attention to three key findings

[11] We focus in this chapter on the questions concerning "overall" satisfaction or happiness with life. Many surveys include in addition questions on satisfaction with specific life domains, such as health or job market status.

[12] A more complete discussion of the issue can be found in Chapter 4 of Fleurbaey and Blanchet (2013).

that are relevant to our quest for an attractive measure of individual well-being. First, the answers to the subjective well-being questions are empirically robust and show regular patterns that are intuitively reasonable. This is not a trivial finding in the light of the traditionally dominant view that interpersonal comparisons of utility are necessarily normative and can have no empirical basis (Robbins, 1938).

Second, the literature convincingly shows that answers to subjective well-being questions are not only, perhaps not even mainly, determined by monetary income or material consumption. The life dimensions that have been found to be relevant for subjective well-being almost coincide with the lists of functionings that were described in the previous section: in addition to income, also health, job market status, quality of relations and social interactions, and even political rights and freedom of speech have been shown to have a significant impact (Frey and Stutzer, 2002). The initial interest in subjective well-being in economics has largely been driven by the striking findings of Easterlin (1974), showing that despite the strong economic growth since the Second World War, subjective well-being has remained almost constant. Recent work (e.g., by Stevenson and Wolfers, 2008) has questioned the empirical validity of the so-called Easterlin paradox, but has not shaken the consensus that subjective well-being is crucially influenced by nonmaterial factors.

A third finding is also related to the Easterlin paradox. Respondents rate their subjective well-being by comparing their actual situation with a set of variable reference standards. They compare their own life conditions with those of their reference groups. Moreover, there is a dynamic process of adaptation of standards through which people lower their aspirations when things go badly and raise their standards when things go well. The empirical literature on subjective well-being now offers plenty of examples showing that adaptation is indeed a pervasive real-world phenomenon. The most striking examples of adaptation are found in the sphere of health, but they also occur in other dimensions of life. Deaton (2008) finds that countries with higher rates of HIV prevalence do not systematically report a lower life (or even health) satisfaction, whereas individuals (and countries) care about HIV and would prefer not to suffer from it. Individuals who have lost a limb may, after adaptation, recover a good subjective well-being score—but still express a strong aversion to disability (Loewenstein and Ubel, 2008; Oswald and Powdthavee, 2008). Interestingly, the recent work on subjective well-being has produced convincing empirical evidence confirming Sen's concern about physical-condition neglect.

The literature on subjective well-being is largely descriptive and seldom engages in an explicit normative discussion. Yet, even if there are no explicitly normative conclusions drawn, often it seems implicitly understood that a higher subjective well-being is better. Using subjective well-being as a measure of individual well-being is understandable given the dominance of subjective welfarism in (applied) economics. A successful measurement of utility allows shortcutting the use of its imperfect monetary approximations (such as the consumer surplus). Moreover, the data on life satisfaction are readily available and easy to collect.[13] If one is willing to accept the answers as an interpersonally comparable measure of utility, one obtains a ready-to-use one-dimensional measure of well-being expressed on a convenient scale, which can then be plugged into a social welfare function. All relevant nonmonetary dimensions are included in the measure, based on the personal evaluation of the individuals themselves. Relying on the earlier notation, this approach uses the subjective well-being scores as aggregator of the various life dimensions and therefore as the measure of individual well-being:

$$WB^{SA}(\ell_i, R_i, S_i) = S_i(\ell_i), \tag{2.2}$$

where the superscript SA refers to satisfaction.

Using subjective well-being as the measure of well-being (i.e., returning to subjective welfarism) implies that it is ethically desirable to redistribute from someone with a high level of subjective well-being to someone with a low level of subjective well-being. It is regrettable that the happiness and welfare economic literature have developed largely in separation, so that there is little debate in the former literature on Sen's original—but very topical—arguments of "physical-condition neglect" and "valuation neglect" against this subjective welfarist position (see also Section 2.2). This separation may partly be caused by a difference in focus because the happiness literature is more interested in "average" or "aggregate" results at the level of the whole society and less on inequality and redistribution. Yet, this different focus offers only a partial explanation, as there is by now a growing number of papers focusing on inequality in happiness (see Dutta and Foster, 2013; Van Praag and Ferrer-i-Carbonell, 2009, for instance).

Before turning to the central questions of this chapter about respect for preferences and responsibility, we first have to return to the question of whether happiness and life satisfaction indeed measure a single concept of utility.

2.3.2.1 Feelings of Happiness and Hedonic Welfarism

The conclusion that there is one concept of utility, underlying both happiness and life satisfaction measures, is contested by most psychologists. They emphasize that "well-being" is a multifaceted experience, and that it is at least essential to distinguish two of its components: affects (feelings, emotions) and cognitions (Diener et al., 1999; Kahneman and Krueger, 2006). For the cognitive component, individuals take some distance to formulate a judgment over their lives. Positive and negative emotions, on the other hand, come in a permanent flow when individuals are awake. They are related to Bentham's pleasures and pains. If one accepts this distinction, the finding that the answers to questions on happiness and life satisfaction are highly correlated becomes worrying rather than reassuring. It suggests that what is measured by the questionnaires is a kind of hybrid mixture of feelings and cognitions without much psychological relevance.

[13] Pragmatic arguments are emphasized, e.g., by Oswald (1997).

In fact, this is precisely the judgment of psychologists like Kahneman, who argue that to measure the affective experience of happiness, other methods (such as experience sampling or day reconstruction) should be used (Kahneman and Krueger, 2006). Application of these methods shows that adaptation is even stronger for feelings than for judgments of satisfaction. Individuals seem to be characterized by a (largely genetically determined) baseline level of happiness, to which they return after having experienced positive or negative shocks.

The distinction between affects and cognitions is not only psychologically relevant but also resonates with welfare economic arguments. It is common to distinguish two variants of welfarism. "Hedonic welfarism" bases the evaluation of individual well-being on feelings of happiness; "preference welfarism" starts from judgments about what is a valuable life and aims at respecting these preferences. There is a clear link with the distinction between affects and cognitions.

Let us first comment on "hedonic welfarism," a modern version of traditional Benthamite utilitarianism. Influential advocates of this variant of welfarism are Kahneman et al. (1997, 2004) and Layard (2005). This approach argues that only feelings of happiness matter for well-being.[14] One of the main reasons for adopting hedonic welfarism is skepticism about the idea that preferences over life dimensions can be meaningfully defined. This skepticism toward preference welfarism leaves hedonic welfarism as the only feasible approach if one cares about individual well-being as experienced by the individuals themselves. Yet, it seems a quite radical position to state that human beings have no single idea about what is valuable in their lives.[15] Rejecting this extreme position, we turn to the central questions of this chapter on respect for preferences and responsibility.

Defining individual well-being in terms of feelings of happiness alone does *not* respect preferences. Such feelings may be very important to individuals, but they are not the only consideration entering the assessment of life (Benjamin et al., 2012; Nussbaum, 2008). Individuals may consider Vincent Van Gogh's life to be more valuable than that of another person who had only pleasant feelings but did not leave any trace after his death. Moreover, using feelings of happiness as the measure of well-being for inequality measurement implies that individuals are not to be held responsible for any factor that influences these (after all extremely subjective) feelings. One immediately runs into the issue of expensive tastes, which is nicely illustrated by Sen's colorful story of the unhappy millionaire:

[14] Layard (2005), p. 121) writes: "Ethical theory should focus on what people feel, rather than what other people think is good for them."

[15] One can easily admit that the preference relation R_i is incomplete, or that there may be inconsistencies in individual evaluations of what is a good life. We will come back to these issues in the next section.

> *I haven't seen you for many years-since I was chucked out of school in fact. I run into you one day in the West End waving at me from your chauffeur-driven Rolls-Royce, looking shockingly prosperous and well-heeled. You give me a ride, and invite me to visit you at your mansion in Chelsea. I remark that I am pleased to see what a high standard of living you are enjoying. 'Not at all', you reply, 'My standard of living is very low. I am a very unhappy man.' 'Why so?' I have to probe. 'Because', you reply, 'I write poems - damn good ones too - but nobody likes my poems, not even my wife. I am always depressed about this injustice, and also sorry that the world has such deplorable taste. I am miserable and have a very low standard of living.' By now I can see no reason to doubt that you are indeed unhappy, but I feel obliged to tell you that you don't know the meaning of 'standard of living'. So you drop me off at the next Tube station (remarking: 'My standard of living high/What a plebeian lie!', adding to the set of people who don't think much of your poetry).*
>
> ***Sen (1984, p. 75)***

Arguably, it is not ethically attractive to compensate the unhappy millionaire for his lower level of subjective well-being. As a matter of fact, given the strong adaptation of happiness feelings, it is even unlikely that any redistribution of income would contribute to a higher level of social welfare. Hedonic welfarists therefore emphasize the importance of investing more in the mental health of the citizens (see, e.g., Layard, 2005)—to the point perhaps of convincing the unhappy millionaire that he is on the wrong track.

While "hedonic welfarism" reduces the scope of individual well-being to feelings of happiness, a broader scope on well-being does not necessarily discard these feelings completely. Indeed, "it would be odd to claim that a person broken down by pain and misery is doing very well" (Sen, 1985, p. 17). It is easy to integrate this intuition in other approaches to well-being, by treating emotions as aspects of life over which individuals may have preferences. In our formal notation, they are then seen as one (but definitely *not* the only) component of vector ℓ_i.[16] In this interpretation, hedonic welfarism respects preferences only under the unrealistic assumption that the only thing that individuals ultimately care about is their own feelings (i.e., a subset of ℓ_i). Yet, including feelings in the list of dimensions of life raises some additional hard questions. Here are some examples. The CEO of a large firm may "need" a certain material lifestyle to be respected in his group of peers, whereas a university professor in a philosophy or welfare economics department may perhaps earn more prestige through a sober lifestyle (Robeyns, 2006). Do we accept these "needs" in our definition of well-being? And, what about feelings of depression that are not obviously linked to physical conditions? Where should we draw the line between real psychiatric problems (which most observers would include in the definition of well-being) and overly subjective reactions, which can be easily manipulated and are situated within the sphere of private information? Leaving these questions aside, we now move to "preference welfarism."

[16] See, among others, Kimball and Willis (2006), Loewenstein and Ubel (2008), and Rayo and Becker (2007).

2.3.2.2 Life Satisfaction, Experienced and Decision Utility

Rather than interpreting subjective well-being as an expression of emotions, one can also see it as reflecting a cognitive judgment about the extent to which one is leading a good life. The satisfaction function S_i is then basically an aggregation function, giving a synthetic evaluation of the complete vector ℓ_i of relevant life dimensions (as we have just seen, this vector may include some pleasant and unpleasant feelings). Various authors in the literature argue that using the satisfaction function means that one evaluates well-being with the value system that is used by the respondents themselves. "If we accept the Marxist idea of 'false consciousness,' we play God and decide what is good for others, even if they will never feel it to be so" (Layard, 2005, p. 121). The argument seems straightforward: If we care for what people care about, we should care for their own perception of life satisfaction. Even if this reasoning may seem convincing at first sight, it requires some further scrutiny.

To do so, let us first consider the relation between the satisfaction function S_i and the preference relation R_i. Clearly, the idea of respecting preferences is only meaningful if one accepts that individuals have well-considered ideas about the good life that can be represented by a (possibly incomplete) preference relation R_i. The precise interpretation of R_i is somewhat ambiguous in the literature, however. Happiness researchers have introduced a distinction between "experienced utility" and "decision utility" (Kahneman and Sugden, 2005; Kahneman et al., 1997). Although decision utility is linked to prospective choices, experienced utility would be better reflected in survey questions that are answered ex post. It turns out that there is frequently a discrepancy between experienced utility and decision utility, in the sense that individuals apparently misperceive the effects of their choices on their future experienced utility (see, e.g., Gruber and Mullainathan, 2005; Layard, 2005; Stutzer and Frey, 2008). In such cases, it is argued, the focus should be on experienced utility.

It is possible to interpret the limitations of decision utility in two different ways. The first interpretation is that stable preferences do not exist. Preference welfarism then is simply not meaningful, and we are back in the hedonic welfarist approach of the previous subsection, albeit possibly with some (undesirable) confusion between affects and cognitions. The second interpretation is that the relevant preferences about life dimensions should involve correct information and proper deliberation and that they therefore are not always revealed in actual choice behavior—and hence in decision utility. The difference between decision utility and experienced utility then justifies skepticism about the use of *revealed* preferences as a criterion for evaluating well-being and does not exclude the interpretation of S_i as a representation of the true underlying preferences of individuals. As we have stressed already before, R_i does not necessarily coincide with revealed preferences in our formal framework. In fact, if psychological feelings of happiness are part of ℓ_i, and insofar as decision utility suffers from imperfect forecast of the psychological effects of choices, we suggest that R_i should not be equated to decision utility, but should be corrected for mistakes and misperceptions.

With this caveat in mind, we can now put forward the obvious point that a necessary condition for S_i to respect preferences is that it is a representation of the preference ordering R_i:

Consistency Assumption $S_i(\ell_i) \geq S_i(\ell_i')$ *if and only if* $\ell_i R_i \ell_i'$.

The formal analogy between the satisfaction function S_i and the valuation function v_i that was introduced before is obvious. Yet, the interpretation of satisfaction as a cardinally measurable and interpersonally comparable variable, which is common in the applied literature, implies that the satisfaction function is more than a general representation of ordinal preferences. Satisfaction scores select one particular "utility" function from the set of all positive monotonic transformations representing the same preference ordering.

Selecting a particular cardinalization imposes a particular scaling. This scaling will reflect comparisons with reference situations such as the worst possible and the best possible situation, the situation one expected at some earlier stage in life (aspirations or expectations), the situation of one's parents, the situation of reference groups such as one's peers or one's fellow citizens. We can call all such components of the judgment the scaling factors. Obviously, scaling factors may differ across individuals and change over time, as aspirations and the choice of reference groups may change. Moreover, the specific scaling may depend on the way the satisfaction question is formulated, or even on its location in the overall questionnaire. It is crucial to realize that the consistency assumption only refers to ordinal preferences and not to these scaling factors. This immediately implies that preferences R_i, which, as discussed before, do not necessarily coincide with revealed preferences, do not necessarily correspond to experienced utility either. Experienced utility as it is typically observed with empirical methods may incorporate a shift in scaling factors and therefore a change from an initial function S_i to another function S_i'. The consistency assumption says nothing about inequalities of the form $S_i(\ell_i) \geq S_i'(\ell_i')$. We will explore the implications of this insight in the following subsection.

It is hard to test the consistency assumption empirically because in practice it seems almost impossible to make sure that preferences and scaling factors remain fixed when an individual moves from ℓ_i to ℓ_i'. We propose to interpret the assumption as a requirement on the measurement of $S_i(\ell_i)$. In other words, we assume that the satisfaction question is sufficiently well crafted so that the answers reflect the individual's views about what is good in life, as embodied in R_i. This is not an innocuous assumption,[17] but from now onward, we will accept that it holds, as it is a necessary condition for subjective well-being measures to respect preferences.

[17] See Fleurbaey and Blanchet (2013) for a discussion. The authors discuss problems related to *scope* (what part of ℓ_i is relevant), *ranking* (how does ℓ_i stand in the set of relevant possible lives), and *calibration* (how does a position in the ranking translate into a category of the questionnaire).

2.3.2.3 Respect for Preferences

Using life satisfaction as a measure of individual well-being does respect preferences, provided that the consistency assumption holds. Indeed, combining the latter assumption with expression (2) immediately confirms that a measure of well-being directly based on $S_i(\ell_i)$ satisfies the personal-preference principle that was introduced in the previous section. Note that this necessarily implies that such a measure will not satisfy the dominance principle.

However, one can argue that respecting preferences for the measurement of well-being requires going beyond intrapersonal comparisons at one point in time (this is the scope of the personal-preference principle). For evaluating inequality, the idea of respecting preferences needs to be extended to situations where scaling factors are different. Such cases can reflect interpersonal comparisons between a pair of individuals who share the same preferences but have different scaling factors, or comparisons over time for a given individual with stable preferences and shifting scaling factors. Consider two triplets (ℓ_i, R_i, S_i) and $(\ell_i' \ R_i, S_i')$ such that $\ell_i \ P_i, \ \ell_i'$ but $S_i(\ell_i) \le S_i' \ (\ell_i')$. There is a preference for ℓ_i against ℓ_i', but the situation (ℓ_i', R_i, S_i') exhibits a greater or equal level of satisfaction. This configuration is not a mere theoretical curiosum. The empirical happiness literature contains many examples of shifting scaling factors. Recall the earlier example of the individuals who express a preference for not being disabled but, after having lost their limbs, recover a good satisfaction score because their aspirations have been adapted to their actual situation (Loewenstein and Ubel, 2008; Oswald and Powdthavee, 2008). Graham (2009) insists that the diversity of scaling factors across individuals generates "happy peasants and miserable millionaires." Her findings do not imply that the poor would prefer to remain poor above getting rich, neither that the rich would prefer to be thrown into poverty. All these examples can be understood as cases of shifts or differences in scaling factors with common preferences in the background. In these examples there is a clear echo of Sen's criticism of "physical-condition neglect" toward subjective welfarism.

It is therefore not obvious that using life satisfaction as a measure of well-being does indeed respect preferences. If one endorses the value judgment that the happy poor are worse off than unhappy millionaires, and hence that redistribution from the millionaire to the poor would lead to an improvement from a social welfare point of view, one has to give priority to the information about (common) preferences over the information about satisfaction levels. This idea is embodied in the following principle, which logically strengthens the personal-preference principle:

Same-Preference Principle: (ℓ, R, S) *is at least as good as* (ℓ', R, S') *if* $\ell R \ell'$, *and strictly better if* $\ell P \ell'$.

Clearly, the same-preference principle is not satisfied by an approach that uses life satisfaction scores as individual well-being measures.

2.3.2.4 Responsibility and Freedom

The discussion in the previous subsection also addresses our second central question on responsibility and freedom. Using life satisfaction as the measure of well-being and as the equalizandum in egalitarian policy implies that redistributing from i to j is ethically desirable if $S_i(\ell_i) > S_j(\ell_j)$. We have seen, however, that in this case it is possible that both individuals prefer ℓ_j to ℓ_i, so that the difference between the life satisfaction scores would only reflect a difference in scaling factors such as aspirations. By choosing this well-being measure, individuals are not held responsible for their aspirations and are compensated for them. Redistributing from a poor peasant to a rich millionaire would then be ethically desirable if the rich millionaire were less satisfied with life because the millionaire would not be held responsible for his ambitious aspirations.

As another example, consider two individuals occupying the same job. The first individual comes from a poor family and has received little education: he is satisfied to have found a job. The other individual has rich parents and a university degree: he is dissatisfied because he is convinced that he was entitled to a "better" job.[18] Because using life satisfaction scores as the measure of well-being does not question the higher aspirations of the rich person, some redistribution of income from the modest to the ambitious individual is socially desirable. This is a conclusion that will appear counterintuitive to many.

2.3.3 Respecting Preferences: Equivalent Income

We have seen that neither the capability approach nor the happiness approach (in its hedonic or its satisfaction interpretation) respects the same-preference principle. The third approach covered in our survey, that of equivalent income or money-metric utility, does.[19] The somewhat surprising history of the concept has been sketched in Section 2.2. We will first introduce the approach and then turn to the most prominent points of criticism.

2.3.3.1 The Equivalent Income

Let us write the vector of relevant life dimensions for individual i as $\ell_i = (y_i, x_i)$, where x_i contains all the nonincome dimensions and y_i his income.[20] Now, choose reference values $\widetilde{x}$ for all the nonincome dimensions. The choice of reference values is a crucial question, to which we will return in the following subsection, but let us first assume that

[18] The empirical relevance of this example is supported by data on Belgian school-leavers in Schokkaert et al. (2011).

[19] The equivalence approach, as it has been introduced in the recent welfare economic literature, is broader than the concept of equivalent income, on which we focus in this chapter (see Fleurbaey and Blanchet, 2013; Fleurbaey and Maniquet, 2011). First, the choice of equivalent income with reference values for all nonincome dimensions is a special case of an approach in which the well-being levels of individuals are ranked on the basis of the intersections of their indifference curves with any monotone path (see, e.g., Fleurbaey et al., 2009). Second, the model can even be further extended to include the notion of equivalent *sets*. In this chapter, we will not elaborate on these generalizations.

[20] In the original literature on money-metric utility, the focus was on comparing consumption bundles, and the nonincome variables then referred to the price vector $\widetilde{p}$. This is just a special case of our approach. Indeed, the vector x may contain the prices of the commodities as one feature of the environment of the individuals.

we can take them as given. The equivalent income y_i^* for individual i is then defined as the solution to the equation

$$(y_i, x_i) I_i (y_i^*, \widetilde{x}). \tag{2.3}$$

In other words, the equivalent income is the level of income that would make the individual indifferent (as judged by his own preferences) between his current situation and the hypothetical reference situation where he would be at the reference values for all nonincome dimensions of life. We then take this equivalent income as the measure of individual well-being:

$$\mathrm{WB}^{\mathrm{EI}}((y_i, x_i), R_i, S_i) = y_i^*.$$

The function that gives the equivalent income for individual i for each combination of (y_i, x_i) is the so-called equivalent income function $y_i^*(y_i, x_i)$.[21]

The concept is illustrated in Figure 2.3 for the case of income–health combinations. Suppose we have to compare the situation of Ann (in A) and Bob (in B). Taking normal health as the reference for the health dimension (we will see in the next section that this is indeed an attractive choice), we can define the equivalent income y_A^* for Ann as the income that would bring her in situation A' (i.e., the income, normal health bundle) that is just as good for her as her actual bundle A. Similarly, we obtain for Bob an equivalent income of y_B^*. We see that Bob's well-being (in B), as measured by y_B^*, is larger than Ann's well-being y_A^* (in A). Conveniently, the equivalent income is expressed in monetary terms. It has all the operational advantages of a cardinal and interpersonally comparable measure, so that it can be used in traditional inequality measures. On the other hand, however, it takes into account all the relevant dimensions in the vector ℓ_i, weighted according to the preferences of individual i herself.

To see this, note first that the equivalent income is a representation of the preference ordering. Indeed if preferences are monotonic in income, it follows immediately that

$$\ell_i R_i \ell_i' \Leftrightarrow y_i^* \geq y_i^{*'}.$$

This shows that, just like subjective well-being under the consistency assumption, the equivalent income function is one possible "utility" function from the set of all positive monotonic transformations representing the same preference ordering. Contrary to life satisfaction scores, however, this specific cardinalization of the utility function does respect the same-preference principle. This is immediately clear from Figure 2.3. The equivalent income uses only ordinal information about the shape of the indifference curves and is not sensitive to differences in aspirations or expectations captured by the satisfaction function S.

[21] As can be seen from expression (3), the equivalent income depends on the choice for the reference value $\widetilde{x}$. To avoid notational clutter, we suppress this dependency in the notation, however.

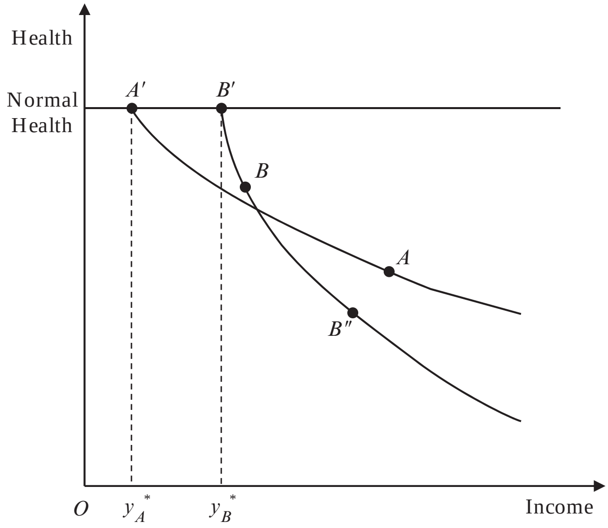


Figure 2.3 Equivalent income.

Recall that a measure that satisfies the personal-preference principle (and, a fortiori, the same-preference principle) does not satisfy the dominance principle. If, in Figure 2.3, Bob were in situation B'' rather than in B, his equivalent income would not change (because B and B'' are on the same indifference curve). He would still be better off than Ann, whereas B'' is dominated by Ann's situation A. The figure also shows why this result is obtained: With his indifference curves, which are "steeper," Bob gives a smaller weight to health than Ann and therefore suffers less from the fact that his health is not at the reference value.

Another way of interpreting the equivalent income refers to willingness-to-pay. It follows from expression (3) that

$$y_i^* = y_i - \mathrm{WTP}_i(x_i \to \tilde{x}; y_i, x_i), \tag{2.4}$$

where $\mathrm{WTP}_i(x_i \to \tilde{x}; y_i, x_i)$ denotes the willingness-to-pay of individual i for a move from x_i to $\tilde{x}$. It is clearly conditional on the level of the actual income level y_i and the level of other life dimensions contained in x_i. Because this willingness-to-pay can be large, it is obvious from expression (4) that the ranking of individuals on the basis of equivalent incomes can be very different from the ranking on the basis of their incomes. This is also illustrated in Figure 2.3, where Bob obtains a larger equivalent income than Ann, despite the fact that his income is smaller.

2.3.3.2 Concavity Failures and Choice of Reference Values

One of the reasons why the equivalent income approach lost popularity in the applied welfare economic literature was the finding by Blackorby and Donaldson (1988) that the equivalent income function $y_i^*(y_i, x_i)$ is not necessarily concave in income. As expression (4) shows, this will occur if $\partial^2 \mathrm{WTP}_i / \partial y_i^2 < 0$. In general, the problem is avoided if preferences are homothetic. Blackorby and Donaldson (1988) then argued that using y_i^* as an indicator of individual well-being in a social welfare function may lead to undesirable redistributive consequences. A regressive income transfer (i.e., a transfer of income from someone with a low equivalent income to someone with a high equivalent income) might be seen as a welfare improvement. Of course, this point only concerns income transfers: A "transfer" of equivalent income itself will lead to an increase in the value of any social welfare function that is concave in equivalent incomes (and to a decrease in inequality with any inequality measure satisfying the Pigou–Dalton transfer principle).

A similar result has been found in the theory of fair allocation, stating that any approach—and not only an approach based on the equivalent income function—that evaluates well-being on the basis of individual indifference curves may clash with a multidimensional Pigou–Dalton transfer principle (see Fleurbaey and Maniquet, 2011). We return to this issue in Section 2.4, but note here already that the only way to avoid the problem is to work with a social welfare function that is of the leximin type (i.e., gives absolute priority to the worse off).

A second point of criticism relates to the dependency of the equivalent income method on the choice of the reference parameters $\tilde{x}$. The dependency is clear: if one moved the reference line in Figure 2.3 sufficiently downward, the relative well-being positions of Ann and Bob would change. Yet, the fact that reference values have to be chosen does not mean that they are necessarily arbitrary (which is the position taken by Blackorby and Donaldson, 1988 and Donaldson, 1992). Given that we are looking for an answer to the normative question "equality of what?" normative choices are inevitable. It is then better to make them explicit, so that they are open to debate and scrutiny. This is precisely the approach taken in the literature on fair allocation. Fleurbaey and Maniquet (2011) provide many examples of applications of money-metric utility in which the reference is selected on the basis of clear normative principles. We will focus here on the choice of references for the calculation of equivalent incomes.

Suppose we want to compare the well-being of two individuals who are in the reference situation for all the nonincome dimensions (e.g., in Figure 2.3 we would consider Alice in A' and Bert in B'). By definition, their equivalent income then coincides with their actual income—implying that the comparison of well-being levels for Alice and Bert reduces to a comparison of their actual income levels, despite the fact that they have different preferences (Bert cares less about health). This example suggests a general criterion for choosing the reference situation. Reference values should be set in such a way that we can accept the implication that when all individuals are in the reference situation for the nonincome dimensions, differences in preferences do not matter to determine who is worse or better off. Namely, if all individuals are in the reference situation, we can focus on income only.

When there exists a "normal" level for the nonincome dimensions to which all individuals aspire, it seems natural to take this normal value as the reference. An obvious example is health because we may assume that, despite some interpersonal differences, there is a large degree of consensus about what is a normal, unproblematic health level. Return to the example of Alice and Bert. It appears counterintuitive to claim that a Pigou–Dalton transfer of income from Bert to Alice would lead to a more *un*equal distribution of well-being on the grounds that Bert cares less about health because they are both in normal health anyway. On the contrary, redistributing income may increase inequality when comparing two individuals at the same health level, when this health level is not the normal one. Indeed, it may happen that the richer individual cares more about health and therefore suffers more from this health condition than the poorer individual. Recall that a similar reasoning has already led us to the conclusion that deviations from the dominance principle can be justified when preferences differ. With this choice of reference, the equivalent income measures the welfare loss that results from deviations from the "normal" level, and this loss is dependent on preferences, which can also be seen from expression (4).

However, it is not possible to define a normal level to which everybody would aspire, for all life dimensions. First, the idea of a "normal" level may be different for different individuals. Leisure (or hours worked) offers a challenging example. Although it may be safe to assume that employment is desirable for everybody (and hence that being employed is a good choice for the reference), people are likely to have different ideas about what is a normal amount of hours of work (and hence of leisure time). Some individuals (academics?) love their work. Others only have access to unpleasant jobs—and although they prefer having a job to being unemployed, they would prefer to have to work as few hours as possible. As shown in Fleurbaey and Blanchet (2013), such differences can be accommodated by selecting individual-specific "normal" values as the reference. Although this complicates the calculations, the interpretation still holds that the difference between income and equivalent income is the welfare loss that results from deviations from the normal level.

Second, even this personalized approach does not work well when the nonincome dimension is unbounded and people have monotonic preferences over it. It is not interesting in this case to take the "best" or a very large reference value because this would lead to extremely small equivalent incomes. A practical solution is then to pick some upper bound (or a variable such as the median) as the reference, but this remains rather ad hoc. More theoretical work is needed to solve this issue.[22]

[22] Difficult questions also arise if the commodity prices are part of the vector x. Fleurbaey and Blanchet (2013, Appendix A) suggest that it would be good to have a reference situation that is as close as possible to the actual market situation faced by the individuals. This intuition is indeed closely related to that of a "normal" value. They then suggest to take as a reference the so-called Scitovsky reference price, defined as the supporting price of the bundle λX (where X is the total actual consumption vector) that belongs to the lower boundary of the Scitovsky set (which contains the vectors of total consumption that can be distributed to keep all individuals on their current indifference curve). Although they show that there are good reasons to pick this reference, the normative justification may appear less compelling than in the cases of health or (un)employment.

2.3.3.3 Freedom and Responsibility

As was emphasized earlier, the most significant feature of the equivalent income approach is that it satisfies the same-preference principle. It is important to note that this respect for preferences is in line with the perspective on personal responsibility that has been put forward by some prominent social philosophers. Rawls (1971, 1982) argued that treating persons as autonomous moral agents necessarily implies that they should assume responsibility for their goals and their conception of the good life. Dworkin (1981a,b, 2000) stressed that individuals should be held responsible for their preferences: In his view, an individual cannot sensibly identify with his own preferences about how to conduct his life and at the same time request compensation on the grounds that his preferences are a sort of handicap.

This view on responsibility for preferences is not beyond criticism, however. In the literature on equality of opportunity and responsibility-sensitive egalitarianism, it has been attacked by authors such as Arneson (1989), Cohen (1989), and Roemer (1998). They claim that preferences are often the product of upbringing and social influences, for which individuals cannot be held responsible, and they instead advocate the "common sense" view that individuals should be held responsible only for what they have genuinely chosen. This, however, raises similar questions as the ones that were encountered earlier when discussing the opportunity-set approach to capabilities. Choices are also determined by factors that are not under the control of the individuals. An attractive theory of responsibility as choice seems to require that one corrects for interpersonal differences in the environment and also for differences in the choice-making abilities of the individuals. Yet, this brings us on a slippery slope. Is there any room left for individual responsibility in a deterministic world, if we better and better understand and explain behavior? The question is especially acute within the paradigm of rational choice (Fleurbaey, 2008). In this paradigm, genuine choice is an elusive notion, as individual decisions result from a mechanical optimization exercise with a given objective (preferences) and a given set of options (determined by the budget set and possibly additional constraints). On the other hand, the equivalent income approach makes use of the preferences that are one of the essential building blocks of the economic model. Recall, however, that one should be careful about equating "revealed" preferences with the authentic views of the good life that have to be respected.

Rawls (1971) already made the connection between respecting different conceptions of the good life and the notion of real freedom. In the same spirit, Fleurbaey (2008) defends the view that individuals should be put in good conditions of autonomy and freedom so that they can be the masters of their lives and participate fully in social interactions. He argues that respect for freedom implies respect for personal preferences. In this view, a policy that is successful in reducing inequality in well-being, defined as equivalent income, can also be seen as reducing inequalities in the real freedom of the individuals.

2.3.3.4 Measurement of Preferences

Compared to the other methods proposed in this section, the equivalent income approach requires additional information. More specifically, for each individual one needs to know not only his actual situation in terms of the relevant life dimensions, ℓ_i, but also his preferences, R_i. Although recovering this information may be hard, it is not a hopeless task. Nor is it a new problem. Economics has a long tradition in identifying preferences, both for market and for nonmarket goods. Three methods have been proposed and applied in the literature on equivalent income, each of them with its own strengths and weaknesses.

2.3.3.4.1 Revealed Preferences

The first method uses revealed preferences (i.e., preferences that are derived from an analysis of observed choice behavior). This approach is common in consumption and labor supply analysis. In fact, the first applications of money-metric utilities made use of it. As an example, King (1983) analyzed the welfare implications of housing subsidies with equivalent incomes that are derived from observations of choices on the housing market. Recently, Decoster and Haan (2013) and Bargain et al. (2013) have estimated preferences over consumption–leisure combinations on the basis of a discrete labor supply model. The authors then derived estimates of equivalent incomes for different choices of the reference values.

Within the perspective of the measurement of equivalent income, an important challenge for this approach is to incorporate preference heterogeneity in an adequate way. More generally, the method only works if individuals have a real choice and can be assumed to choose rationally. This observation points at two limitations. First, the revealed preferences approach cannot give information on the relative value of dimensions that are not chosen by the individual. An example is health: Although it can be influenced by lifestyle choices to some extent, health remains largely outside the sphere of private decisions. Second, choice behavior does not always reveal the informed and authentic preferences of individuals. Human beings make mistakes or take decisions under imperfect information or under social pressure. Behavioral economics has shown that it is not always possible to identify preferences in such situations because the outcomes of two different behavioral models (with different underlying preference relations) may be observationally equivalent in terms of choices (Bernheim, 2009; Bernheim and Rangel, 2009).

2.3.3.4.2 Stated Preferences

The second method is based on stated preferences and makes use of the contingent valuation methods that are typically used in environmental economics and in health economics to measure the subjective willingness-to-pay for goods that cannot be bought on a market.[23] The contingent valuation method consists of asking people to evaluate the income they would need to be as well off in different reference scenarios as they are currently. As expression (4) shows, as soon as the actual income of individuals and their willingness-to-pay to be in the reference situation is known, equivalent income can be computed easily. This method is indeed the most direct way to make the notion of equivalent income operational.

To analyze policies, it is generally not sufficient to register the willingness-to-pay and the equivalent income of the individuals. One also would like to derive information about their entire indifference map. Because the equivalent income function $y_i^*(y_i, x_i)$ is a representation of the preference ordering, observations on y_i^* (or on WTP_i) can be used to estimate the parameters of a utility function. Fleurbaey et al. (2013) have used this method to calculate equivalent incomes for income–health combinations with survey data collected in Marseilles.[24] The authors use the estimated parameters to derive a set of distributional weights that could be implemented in a cost-benefit analysis of medical interventions.

It is fair to say that within the economic literature there is no consensus about the validity of these stated preferences techniques. There are strong believers and at the same time ruthless critics. Two titles in a recent symposium of the *Journal of Economic Perspectives* summarize the debate. Carson (2012) claimed that contingent valuation is a practical alternative when prices are not available, whereas Hausman (2012) argued that the results range from dubious to hopeless. This is not the place to settle this debate. Let us simply note that the applications in the context of equivalent income may be among the least contested because they are based on realistic and understandable alternatives, with which the respondents may have had some previous experience (like being in good health), rather than more esoteric alternatives (like the survival of a particular whale) with which they are not familiar.

2.3.3.4.3 Using Satisfaction Data

The third method for estimating preferences makes use of the answers to a "satisfaction with life" (or happiness) question. At first sight, this may seem a surprising venture, given our earlier emphasis on the fact that the answers to these satisfaction questions do not satisfy the same-preference principle. Yet, we have seen that the satisfaction function S_i can be interpreted as one utility function representing the preference ordering of individuals, provided the consistency assumption holds. By modeling carefully the effects of aspirations and expectations on subjective well-being, one can retrieve the ordinal information about preferences that is embodied in the satisfaction answers. This method is in

[23] Other stated preferences methods (e.g., discrete choice analysis) could in principle also be useful to estimate preferences.

[24] A similar analysis with data from a representative survey in France is presented in Fleurbaey et al. (2012).

line with a growing body of research that estimates willingness-to-pay for nonmarket goods through their effects on subjective satisfaction (see, e.g., Clark and Oswald, 2002; Van Praag and Ferrer-i-Carbonell, 2007).

Denoting by π_i the individual characteristics of individual i, which are not seen as life dimensions but as factors that do influence life satisfaction (the scaling factors), we can rewrite the satisfaction function as $S(y_i, x_i; \pi_i)$. The equivalent income can then be computed by solving

$$S(y_i, x_i; \pi_i) = S(y_i^*, \widetilde{x}; \pi_i)$$

for y_i^*. If we adopt a log-linear approximation (which is the dominant model in the empirical happiness literature), this yields

$$\ln y_i^* = \ln y_i - \sum_j \left(\frac{\partial S/\partial x_{ij}}{\partial S/\partial \ln y_i} \right) \left(\widetilde{x}_j - x_{ij} \right), \tag{2.5}$$

where the subscript j refers to the different life dimensions. Expression (5) shows that interpersonal variation in the psychological characteristics π_i will only influence the value of the equivalent income if it influences the marginal rates of substitution. Differences in scaling factors that only influence the satisfaction level, without affecting the relative weights given to the different dimensions, will not influence the estimated y_i^*. The satisfaction method has been used to calculate equivalent incomes by Fleurbaey et al. (2009) and by Schokkaert et al. (2011). In both papers it is shown that the ranking of well-being on the basis of subjective satisfaction differs considerably from the ranking of equivalent incomes.

Like the stated preferences approaches, the use of satisfaction data allows for the incorporation of nonchoice dimensions in the evaluation. However, the precise specification of the function $S(y_i, x_i; \pi_i)$ and the identification of the relative effects of x_i and π_i raise difficult issues.[25] Most important, the method rests on the acceptability of the consistency assumption. To be useful, the satisfaction question should be formulated in such a way that it can safely be assumed to capture the respondent's cognitive views on what constitutes a good life.[26]

2.3.3.5 What If Preferences Are Incomplete?

The equivalent income approach, as described until now, rests on the assumption that well-defined individual preferences exist. Many researchers expressed their skepticism

[25] The treatment of education illustrates the problem. Having a good education may be seen as an important dimension of life (it is, e.g., emphasized by Nussbaum, 2000). At the same time, however, education may also have a direct influence on aspirations (e.g., with respect to job characteristics, as in Schokkaert et al., 2011). With the satisfaction approach it is impossible to disentangle the two effects.

[26] See Fleurbaey and Blanchet (2013) on the wording of subjective satisfaction questions.

about the use of preferences. We have already seen that, in the light of the many instances of differences between "decision utility" (the perceived utility on which decisions are based) and "experience utility" (the real after-decision utility), some researchers from the subjective well-being approach suggest focusing on the latter in case of conflict. Their skepticism seems to be supported by the recent findings of behavioral economics that a large number of "behavioral anomalies" make it difficult to interpret individual choice behavior as the maximization of well-defined preferences. We have argued before that the latter point urges a focus not on revealed, but on "authentic" preferences.

Other researchers reject the idea that individuals "authentically" have a complete preference relation over all possible lives. The assumption of a complete preference relation over all possible lives is indeed a strong one, implying that individuals can order states with which they may not be familiar at all. The psychological uncertainty about preferences may be expected to be larger further away from the actual situation. To calculate healthy-equivalent incomes as depicted in Figure 2.3, for instance, one needs nonlocal information on the indifference curve. Is someone who has been chronically ill for a long time (or is handicapped since birth) able to evaluate trade-offs in a situation of normal health? And, even if individuals have clear ideas about what a good life is for them, the available techniques to recover these preferences are still in their infancy and far from perfect. In fact, it is quite likely that different techniques will lead to conflicting results.

If one does not believe that authentic preferences do exist or can be recovered, one could conclude that the equivalent income approach to measuring well-being is not meaningful and that one has to go back to either more "objective" applications of the capability approach or the direct use of subjective satisfaction measures. An alternative approach, however, is to keep individual preferences as the underlying foundation for measuring individual well-being, but to accept that the preference relation is not complete if choices (or stated preferences) are conflicting and context-dependent. This route has been explored by Fleurbaey and Schokkaert (2013) who implement the notion of incomplete preferences (or choices) that was suggested by Bernheim and Rangel (2009) in the context of the measurement of individual well-being. They show that incomplete preferences can be accommodated by introducing upper and lower bounds to equivalent income. Figure 2.4 illustrates the suggested method for the example that was already used in Figure 2.3. Suppose the individual has the income–health combination depicted in Z. Imagine that his preference relation is incomplete: bundles in the region UC are seen as better by him, bundles in the region LC are worse, but bundles in the region NC are noncomparable to bundle Z. This way of modeling preferences embodies the natural assumption that individuals have finer preferences when comparing closer alternatives. Figure 2.4 then immediately shows (using the same assumption about the choice of the reference situation as discussed before) that it is possible to derive an upper limit $y_i^{*\sup}$ and a lower limit $y_i^{*\inf}$ for the equivalent income. One can then argue that individual i is better off than individual j if $y_i^{*\inf} > y_i^{*\sup}$. Arguably, this condition is a

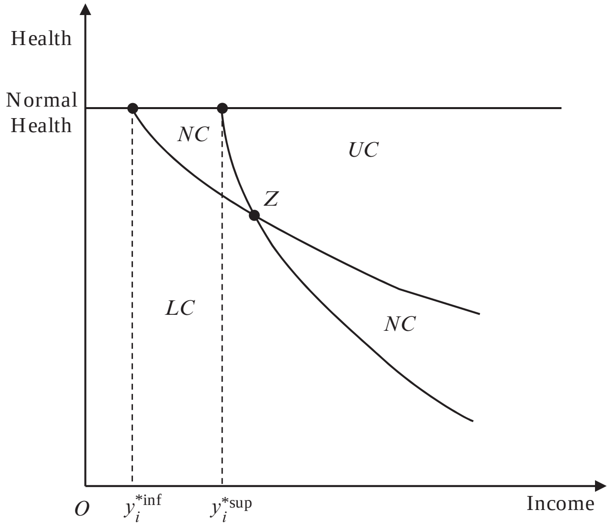


Figure 2.4 Equivalent income with incomplete preferences.

stringent one, and Fleurbaey and Schokkaert (2013) show how it is possible to weaken it to be able to compare more individual situations. The less-stringent conditions do not preclude mistakes in interpersonal comparisons, but they prevent the evaluator from missing a situation in which the worse off is really badly off.[27]

2.3.4 Conclusion

Choosing a meaningful measure of individual well-being for the analysis of inequality is a normative exercise. The underlying value judgment is the following: when is it ethically acceptable to say that one individual is worse off than another, in the sense that redistribution is desirable from a social perspective?

In this section, we focused on the normative foundations of the three most popular approaches for measuring well-being in a multidimensional setting: the capability approach, the subjective well-being approach (with its two variants: hedonic and preference welfarism), and the equivalent income approach. If one interprets the capability approach in terms of (refined) functionings and one opts for a measurement tool that respects preferences, one arrives at the equivalent income approach. Yet, most followers of the capability model are skeptical about respecting preferences and the idea of trading off different life dimensions against each other. In these concluding remarks, we will follow this dominant perspective on the capability approach.

[27] The idea is that if the evaluator is wrong about the worse off in a pairwise comparison, the true worse off is not as badly off as he could be if the mistake was in the opposite direction.

A first ethical requirement for a measure of well-being could be that individual *i* is said to be better off than individual *j*, if he has higher achievements in all life dimensions compared to individual *j*. This is the so-called dominance principle. We have seen that this seemingly innocuous principle cannot be reconciled with respect for preferences and that it is therefore not satisfied by the happiness and equivalent income approaches. It can be satisfied by the capability approach, on the other hand, if the latter is implemented with an aggregation procedure that gives objective weights to all dimensions. This observation immediately implies that the capability approach in this interpretation cannot respect preferences.

Respect for preferences comes in a weak and a strong form. The personal-preference principle is satisfied by the subjective well-being approach, provided that the answers to satisfaction questionnaires are consistent with preferences. The personal-preference principle is also satisfied by the equivalent income approach. Only the equivalent income approach satisfies the stronger same-preference principle, which extends respect for preferences to interpersonal comparisons. The crucial difference between the two approaches is the treatment of aspirations and expectations, and hence of adaptation. These phenomena are taken up in the life satisfaction measure of well-being, whereas they are corrected for by the equivalent income approach. The hedonic version of the subjective well-being approach goes very far in its rejection of the relevance of preferences by putting forward that only ex post feelings should matter to determine who is worse off. This appears to be a radical position, given the widespread observation that people care about more than their subjective feelings.

The choice of a metric of well-being has implications for the implied cut between personal and social responsibility. In its opportunity set interpretation, the capability approach holds people responsible for their choices. This can be harsh in the light of significant differences in the decision-making capacities of individuals. Correcting for these differences is probably easier when we turn to an interpretation in terms of refined functionings. The satisfaction approach compensates individuals with expensive tastes (high aspirations) and will not compensate them if they adapt to poor physical conditions. The equivalent income approach evaluates individual achievements. However, because it evaluates these achievements on the basis of the own conceptions that individuals have about what is a good life, individuals are held responsible for these conceptions.

For practical purposes it is important to realize that the different perspectives on well-being also impose different informational requirements. If one deems that life satisfaction questionnaires yield meaningful answers, this approach is the easiest to implement. The hedonic approach requires that feelings be registered, ideally with a day reconstruction or an experience sampling method. Both the capability and the equivalent income approaches need information about the different life dimensions at the level of the individual. To calculate equivalent incomes one moreover has to know individual preferences. We discussed three methods to retrieve these preferences.

A final word of caution. The happiness literature has looked at the effect of economic inequality on life satisfaction (see also Chapter 13). To give one example, Alesina et al. (2004) show that respondents report a lower level of happiness when inequality is high. This effect is larger and statistically more significant in Europe than in the United States. More strikingly, there are also differences across groups. In Europe mainly the leftist and the poor suffer from inequality, while in the United States the strongest negative effect is on a subgroup of leftist rich. This fascinating result might be explained by differences in perceived mobility. However, for our purposes it is important to be clear about the normative status of these findings. One interpretation is that people care about their social environment, i.e., that perceived inequality (or injustice) is one of the relevant dimensions influencing the quality of their life. This can be easily incorporated in all three approaches in this section. Only the life satisfaction approach would go further, however, and would claim that inequality only matters from an ethical point of view if it influences satisfaction—which would imply that the fact that it is felt as less important in the United States would also imply that it is indeed less important from an ethical perspective. This conclusion is not acceptable for the two other approaches. Both for the capability and for the equivalent income approach, inequality is a problem of justice, and justice remains ethically important, even if people do not (seem to) care.

2.4. MULTIDIMENSIONAL INEQUALITY AND DOMINANCE

The route taken by most welfare economists to evaluate the multidimensional distribution of well-being consists of two steps. In a first step, an appropriate measure of individual well-being is derived by answering the question "equality of what?" In the previous section we studied three prominent answers to that question. In a second step, social welfare or inequality is measured consistently with the analogue of the Pigou–Dalton transfer principle defined in the space of the well-being measures, which have been obtained in the first step. As a consequence, the ethical attractiveness of the Pigou–Dalton transfer principle and the well-being measure are intimately linked.

An alternative, more direct route has been followed in the recent literature on multidimensional inequality. It consists of first generalizing the Pigou–Dalton transfer principle toward a multidimensional framework and then imposing this principle directly in the multidimensional space of achievements. At first sight, this route appears to be shortcutting the problem of constructing a well-being measure in the initial step. We have seen indeed that a number of authors within the capability approach are reluctant to construct one single index of well-being. We will investigate whether the methods developed in the literature on multidimensional inequality allow studying the multidimensional well-being distribution without constructing such an index.

For this section, we introduce some additional information. Consistent with the previous section we assume that a social situation can be described as $(\ell_i, R_i, S_i)_{i=1}^{n}$. In addition, it will turn out to be convenient to summarize only the achievements of all individuals by means of a so-called distribution matrix. Next we give an example of a distribution matrix L of a society with n individuals and m dimensions of life. Let ℓ_i^k be the achievement of individual i in dimension k. As before, ℓ_i refers to the m-dimensional vector of all achievements of individual i (a row of the matrix), and ℓ^k refers to the n-dimensional vector of achievements of all individuals in dimension k (a column of the matrix).

$$L = \begin{array}{cc} \begin{bmatrix} \ell_1^1 & \cdots & \ell_1^m \\ \ell_2^1 & \cdots & \ell_2^m \\ \vdots & \vdots & \vdots \\ \ell_n^1 & \cdots & \ell_n^m \end{bmatrix} & \begin{array}{l} \leftarrow \text{Individual } 1 \\ \leftarrow \text{Individual } 2 \\ \qquad \vdots \\ \leftarrow \text{Individual } n \end{array} \\ \begin{array}{ccc} \uparrow & \cdots & \uparrow \\ \text{Dim. } 1 & \cdots & \text{Dim. } m \end{array} & \end{array} \tag{2.6}$$

The literature on multidimensional inequality studies how to summarize the information in a distribution matrix by means of a single numerical value.[28] By taking a distribution matrix as the only information basis, it is clear that the standard multidimensional social welfare measures proposed in the literature are not sensitive to the preferences held in the society. We return to this topic in the next subsection.

2.4.1 Two-Step Aggregation and Cumulative Deprivation

Although the aggregation of a distribution matrix into a numerical value is not always performed by an explicit two-step procedure, most of the existing multidimensional measures combine two one-dimensional aggregations. One aggregation is across the n individuals in the society. The other aggregation is across the m dimensions of well-being. Different multidimensional measures of social welfare differ in the functional specifications of both aggregations and in the sequencing of both steps.

Let us describe two procedures to sequence this two-step aggregation. In the first procedure, we first aggregate across the different individuals in each dimension. In this step we obtain for each dimension a single summary statistic, so that an m-dimensional vector of summary statistics is generated. In the second step, this vector is further aggregated across dimensions. Kolm (1977) calls this procedure a specific one. Pattanaik et al. (2012) refer to it as the column-first two-step aggregation procedure. In the second

[28] We refer the reader to Chapter 3 or Weymark (2006) for detailed surveys on the literature on multidimensional inequality. Following Kolm (1977), a measure of multidimensional inequality can be derived from a measure of multidimensional social welfare as the fraction of the aggregate amount of each dimension that could be destroyed if every dimension of the matrix were equalized while keeping the resulting matrix socially indifferent to the original matrix. We will focus primarily on measures of social welfare in this section.

procedure, the order of aggregation is reversed: in the first step one aggregates for each individual i the dimensions of well-being, which generates a measure of well-being. All the obtained well-being measures generate an n-dimensional vector of individual well-being measures. In the second step, this vector is aggregated across individuals. Following Kolm (1977) this second procedure will be referred to as an individualistic one, or a row-first aggregation procedure according to Pattanaik et al. (2012).

In general the two procedures lead to different results (see Decancq and Lugo, 2012; Dutta et al., 2003; Kolm, 1977). Most theoretical multidimensional inequality measures follow the individualistic procedure and aggregate first across dimensions and then across individuals. Some authors have followed the other track, however. A notable example is provided by Gajdos and Weymark (2005), who impose separability between dimensions. Imposing this requirement brings them to a specific procedure. Specific procedures have the operational advantage that they allow the use of different information sources for the different dimensions of well-being. The summary statistic of one dimension may come from one survey, whereas the summary statistic of another dimension may be based on a different survey. A prominent example of such an approach is the HDI, which will be discussed in more detail in the next section.

The flexibility of the specific procedure with respect to the data sources comes at a (high) price, however. The second aggregation function used in a specific procedure aggregates across the different dimensionwise summary statistics. This aggregation may appear to be largely arbitrary. Contrary to an aggregation across dimensions of well-being at the individual level, a theoretical framework for aggregation of summary statistics is indeed missing. This arbitrariness probably underlies the reluctance of various researchers and statistical agencies to pursue an aggregation of summary statistics. A portfolio or dashboard of separate summary statistics, which each can be monitored in separation, is often presented as an alternative. This method is consistent with the view that different dimensions of life are incommensurable, as we have encountered, e.g., in our overview of the capability approach.

Irrespective of the choice whether and how the summary statistics are aggregated in its second step, a specific procedure has an additional drawback. An important aspect of the information on well-being is lost, namely the correlation between the positions of the individuals across the different dimensions (see Decancq, 2013, for a discussion). When the dimensions of life are correlated, deprivations in one dimension are cumulated with deprivations in other dimensions. Compare, for instance, the following two distribution matrices L and L':

$$L = \begin{bmatrix} 10 & 10 \\ 20 & 70 \\ 70 & 20 \end{bmatrix} \quad L' = \begin{bmatrix} 10 & 10 \\ 20 & 20 \\ 70 & 70 \end{bmatrix}.$$

In both matrices, there are two dimensions of life (columns) and three individuals (rows). It is easy to see that each of the four dimensionwise distributions is the same and hence that each specific aggregation across individuals should lead to exactly the same result. Yet, in the distribution matrix L', there is one individual who is bottom-ranked in both dimensions of life, another individual who is second-ranked in all dimensions, and still another individual who is top-ranked in all dimensions. This society is arguably more unequal than the society represented by L with exactly the same distributional profile in each dimension, but where the achievements of individuals two and three are more mixed. It seems natural to require that the multidimensional evaluation is at least sensitive to the degree to which deprivations in each dimensions are cumulative across dimensions. Pogge (2002, p. 11), for instance, writes: "Consider institutional schemes under which half the population are poor and half have no access to higher education. We may plausibly judge such an order to be more unjust when the two groups coincide than when they are disjoint (so that no one bears both hardships)."

The preceding example illustrated that all measures obtained through a specific procedure are blind to the correlation between the dimensions of life. It follows that a concern for correlation or an aversion to cumulative deprivation rules out the specific (or column-first) sequencing of both aggregations as well as dashboard approaches.[29] This brings us to the alternative, individualistic sequencing in which the dimensions of life are first aggregated for each individual and then across all individuals. Interestingly, this procedure coincides with the welfare economic approach surveyed in the previous section. Although the literature on multidimensional inequality measurement offers a coherent axiomatic justification for the functional specification of the various measures, the link between the formal axioms used (such as homotheticity or separability) and the normative foundations of the implied well-being measure is usually not explained in detail, however.

2.4.2 Multidimensional Pigou–Dalton Transfer Principles and Respect for Preferences

A central question in the literature on multidimensional inequality deals with the generalizations of the standard one-dimensional Pigou–Dalton Transfer Principle and the restrictions that each of these generalizations impose on the functional specifications of the aggregation across dimensions and individuals.[30] In this section we are particularly concerned with the question whether such generalizations can be reconciled with a general respect for individual preferences.

2.4.2.1 Multidimensional Pigou–Dalton Transfer Principle(s)

In a one-dimensional setting, a Pigou–Dalton transfer consists of transferring a positive amount of income from a richer to a poorer individual without reversing the ranking

[29] See Dardanoni (1995), Gajdos and Weymark (2005), and Pattanaik et al. (2012) for formal discussions.
[30] See Weymark (2006) and Fleurbaey (2006b) for surveys.

between both individuals. A natural generalization into a multidimensional framework is the following (see Fleurbaey, 2006b; Fleurbaey and Maniquet, 2011).

Pigou–Dalton Transfer Principle $(\ell_i, R_i, S_i)_{i=1}^n$ *is strictly better than* $(\ell_i', R_i, S_i)_{i=1}^n$*, if for all individuals* $k \neq i, j$*, we have that* $\ell_k' = \ell_k$*, and for individuals i and j, we have that for* $\delta \in \mathbb{R}_+^m \setminus \{0\}$

$$\ell_i' = \ell_i + \delta \leq \ell_j - \delta = \ell_j'. \tag{2.7}$$

A positive bundle δ is transferred from a donor j to the recipient i, where the donor has achievements that are at least as good as the recipient in all dimensions of life. In the axiomatic literature on multidimensional social welfare, on the other hand, it is more common to work with transfers where the transferred bundle is a fraction of the difference between the achievement vector of donor and recipient of the transfer, so that $\delta = \lambda(\ell_j - \ell_i)$. Moreover, often expression (7) is replaced by the following expression,

$$\ell_i' = \ell_i + \lambda(\ell_j - \ell_i) \text{ and } \ell_j' = \ell_j - \lambda(\ell_j - \ell_i), \tag{2.8}$$

for $\lambda \in (0,1)$.[31] The most important difference between expressions (7) and (8) is the fact that the achievements of the donor of the transfer should no longer be larger than the achievements of the recipient in all life dimensions. Consequently, the transfers may go in opposite directions for different dimensions. Consider the following distribution matrices L and L'' for example, where

$$L = \begin{bmatrix} 10 & 10 \\ 20 & 70 \\ 70 & 20 \end{bmatrix} \quad L'' = \begin{bmatrix} 10 & 10 \\ 50 & 40 \\ 40 & 50 \end{bmatrix}. \tag{2.9}$$

One easily checks that a transfer of 30 units is carried out between individuals 2 and 3 in distribution matrix L to reach matrix L''. In the first dimension the units are transferred from individual 3 to individual 2, whereas in the second dimension the 30 units are transferred in the other direction from individual 2 to individual 3.

This example illustrates a fundamental problem with using expression (8) in a richer setting where individuals may have different preferences (Fleurbaey, 2006b). Distribution matrix L'' is obtained from L by a multidimensional transfer. Yet, individual 2 may prefer his bundle in the distribution L to the one in L'', as he may give more weight to the second dimension. Also individual 3 may prefer his bundle in the distribution L, if

[31] Any sequence of these transfers can be written as a bistochastic matrix (see Weymark, 2006, for details). The converse of this statement does not always hold. When $n \geq 3$ and $m \geq 2$ not all bistochastic matrices can be obtained as a sequence of transfers described by expression (8) (Marshall and Olkin, 1979, p. 431). The class of multidimensional transfers that can be expressed by means of a bistochastic matrix is the workhorse of many (axiomatic) studies of multidimensional inequality. It has the advantage of imposing a clear structure on the functional specifications of the aggregation across dimensions and individuals (see, e.g., Kolm, 1977; Tsui, 1995).

he cares more about his achievement in the first dimension. The transfers may therefore go against unanimous individual opinions on the change in well-being. At first sight, this problem seems to be avoided by restricting the transfers to cases where the donor vector dominates the recipient, as in the definition of a multidimensional Pigou–Dalton transfer based on expression (7), so that there is an unambiguous recipient who benefits from the transfer and an unambiguous donor whose well-being is worsened. Yet, we will now see that even these transfers are incompatible with a respect for preferences.

2.4.2.2 The Impossibility of a Paretian Egalitarian

Let us assume, as in the previous section, that all individuals have an informed judgment about what a good life is. Respect for these individual opinions may in this context be expressed by the following Pareto condition:

Weak Pareto Principle $(\ell_i, R_i, S_i)_{i=1}^n$ *is strictly better than* $(\ell_i', R_i, S_i)_{i=1}^n$ *if for all* $i, \ell_i P_i \ell_i'$

The (weak) Pareto principle and the multidimensional Pigou–Dalton transfer principle conflict as soon as at least two individuals have different preferences. This impossibility result is intuitive. The Pareto principle requires that individual preferences are respected, whereas the multidimensional Pigou–Dalton transfer principle advocates some transfers irrespective of the individual preferences. Figure 2.5 illustrates a simple graphical proof (see Fleurbaey and Maniquet, 2011, Theorem 2.1, and also Fleurbaey and Trannoy, 2003). The Pareto principle requires that distribution matrix L^1 is strictly better than L^4 because for all individuals the achievement vector in L^4 is below the indifference curve containing the achievement vector in distribution matrix L^1. Similarly, L^3 is strictly better than L^2. On the other hand, the multidimensional Pigou–Dalton transfer principle requires that L^2 is strictly better than L^1, and L^4 is strictly better than L^3, which creates a cycle.

This impossibility reflects a deep tension between two ways of interpreting what it means to respect unanimous preferences. As the donor of the transfer has a higher achievement in all dimensions of life than the recipient, all individuals with monotonic preferences will agree that the donor is indeed better off, so that a transfer from the donor to the recipient is a social improvement. On the other hand, it may be the case that all individuals are indifferent between the current distribution and a new one, where the initial donor now has a lower achievement in all dimensions of life than the initial recipient. Note that the same tension is underlying the incompatibility between the personal-preference principle and the dominance principle that was discussed in the previous section.

The impossibility result brings us to a crossroad. We can take two directions from here. Either we give priority to the Pareto principle and look for appropriate weakenings of the multidimensional Pigou–Dalton transfer principle. This route is taken by Fleurbaey and Maniquet (2011), among others. A natural weakening of the multidimensional Pigou–Dalton transfer principle is to impose the additional requirement that both donor and recipient of the transfer should have the same preferences (i.e., agree on the

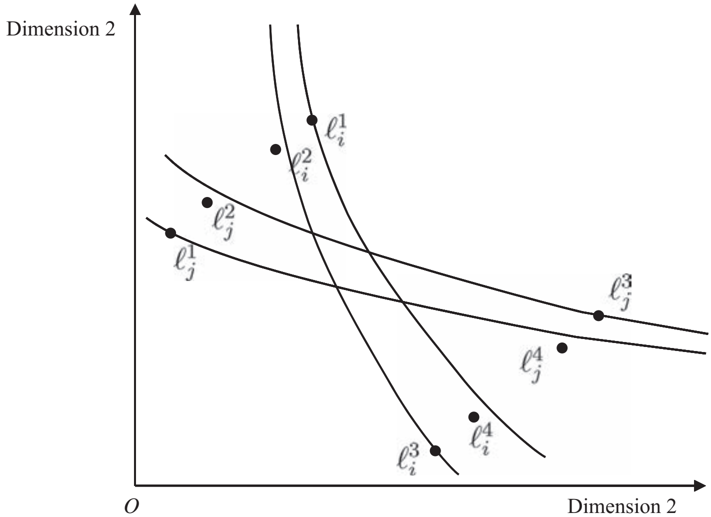


Figure 2.5 Weak Pareto principle and multidimensional Pigou–Dalton principle are incompatible.

good life). This restricted transfer principle is arguably a weak one (as it remains silent on the evaluation of all transfers where donor and recipient disagree on the good life), yet in terms of implications on the social welfare function, it turns out to be very strong. Together with the Pareto principle and the requirement that the comparison of two allocations only depends on the indifference curves at these two allocations, it imposes a leximin aggregation across individuals that gives priority to the worse off. This result echoes our earlier findings when we described the problems with the nonconcavity of the equivalent income in the previous section.

Alternatively, one can give priority to the multidimensional Pigou–Dalton transfer principle. This implies that the resulting social evaluation procedure will not be able to respect individual preferences. The literature on multidimensional inequality measurement has taken this second route by assuming that the well-being of a society can be described using information on achievements alone (i.e., a distribution matrix), disregarding information on the preferences of the individuals themselves.[32] This assumption imposes the requirement that the social welfare function is anonymous in the achievement space (see, for instance, Kolm, 1977; Tsui, 1995; Weymark, 2006). A social welfare function is anonymous in the achievement space whenever permuting individual achievement vectors is a matter of social indifference. As a consequence, the well-being measures used to aggregate across dimensions are identical for all individuals. This assumption is defended as requiring equal treatment of all individuals i with the same achievement bundle ℓ_i either because the observer is unable to distinguish between other

[32] For a more subtle weakening, see Sprumont (2012).

possibly relevant individual characteristics (such as the individual opinions on what constitutes a good life) or because the observer considers the other individual characteristics to be ethically irrelevant.

In his seminal article Kolm (1977) suggested that a common well-being measure can be seen as "the observer's evaluation of the individual welfare." Alternatively, the common objective opinion on the good life is rooted in some "reasoned social agreement on basic components of well-being and on the relative "urgency" of claims to different goods" (Scanlon, 1975). These options are closely linked to the perfectionist approach and the focus on public reasoning that we found within the capability approach by Nussbaum and Sen, respectively.

2.4.3 Dominance and Agnosticism on Preferences

Whether a social agreement on the components of well-being can effectively be reached is doubtful. Yet, even if it is hard to reach a social agreement on which common well-being measure to use, it may be possible to reach an agreement on some of its basic features while remaining agnostic on other features. This agnosticism comes at a price, as the social evaluation criterion will become incomplete and indecisive on the comparison of some social situations.

The dominance approach studies these incomplete orderings. In their seminal contribution, Atkinson and Bourguignon (1982) extended the existing one-dimensional dominance approach to a multidimensional framework.[33] A distribution matrix is said to dominate another one if the sum of well-being measures is greater for each and every well-being measure in a given set of measures that satisfy certain sign-restrictions on its partial derivatives.

In general, the class of measures that satisfy given sign restrictions contains infinitely many members, so that checking for dominance involves checking infinitely many inequalities. Luckily, dominance with respect to classes of well-being measures can be shown to be equivalent with implementable criteria. In a two-dimensional framework, Atkinson and Bourguignon (1982) showed, for instance, that dominance with respect to the class of increasing well-being measures with a negative cross-derivative is equivalent to first-order stochastic dominance in terms of the joint distribution functions corresponding to the distribution matrices. Various statistical tests have been developed to test whether distribution functions first-order stochastically dominate one another (see Chapter 6). By imposing a negative cross-derivative, the marginal increase in well-being from having a small increase in the achievement of the first

[33] Related approaches focus on the measurement of inequality, rather than social welfare. These approaches start from a multidimensional generalization of the Lorenz criterion, based on the so-called zonotope (Koshevoy, 1995). Unfortunately, the equivalence between second-order stochastic dominance and the Lorenz criterion breaks down in the multidimensional case.

dimension decreases with the level of the achievements in the second dimension. In other words, if some manna would become available in the first dimension, the social planner prefers it to go to the worse-off individual in the second dimension. This restriction introduces again some aversion to correlation and cumulative deprivation between the two dimensions of well-being. Atkinson and Bourguignon also looked for the consequences of imposing further restrictions on the partial derivatives, and later work has extended these results (see Trannoy, 2006, and the references therein). The more sign-restrictions are imposed, the more complete the ordering becomes. However, the results become arguably harder and harder to interpret, as higher-order cross-derivatives are involved.

The dominance approach moves us away from the perfectionism that is implicitly underlying approaches that impose a single well-being measure for all individuals. Yet, the unanimous judgment of a class of social welfare functions remains based on a common well-being function for all individuals, so that the dominance approach ignores the diversity of individual preferences. Whether one finds this problematic or not depends on the attitude one takes toward the idea of respecting preferences. Multidimensional inequality measures and dominance approaches are arguably the best way to proceed if one believes that individuals do not have well-defined conceptions of the good life, or that, even when they exist, it is impossible to know them, or that, even when they exist and one can approximate them, one should not do so but rather implement an objective conception of the good life. Again, this is an essentially normative debate.

2.5. APPLICATIONS

Although our discussion so far has remained at an abstract level, the different positions described suggest different approaches to many applied issues that are of great importance for measuring inequality. An important application is the booming literature on socioeconomic inequality and racial disparities in health, in which the issue of cumulative deprivation (with respect to income and health) plays a crucial role. Because this literature has been discussed in great detail in Fleurbaey and Schokkaert (2012), we will not repeat this analysis here. We will illustrate the practical relevance of the previous sections by focusing on three applications. We first discuss the issue of household equivalence scales and (related to that) the measurement of intrahousehold inequality. We will then look at the different methods that have been proposed to include the value of public goods and services into the analysis of inequality. Our third application is the analysis of world inequalities, including a discussion of purchasing-power parity (PPP) indices. In each of these subsections, we do not go into the technical details but focus on the relationship with the normative analysis in the previous sections.

2.5.1 Household Equivalence Scales

It is widely agreed that the quality of social relations is one of the most important dimensions of life. For many people this is particularly true for their relationship with a partner and the quality of their family life. Also the presence of children changes life deeply (for better or for worse). Therefore, it seems natural to include these family-related dimensions in a broader view of well-being. Family relations have been introduced in the capability approach, often with a focus on gender issues (see, for instance, Nussbaum, 2000; Robeyns, 2003). Moreover, family relations have been shown to have a strong effect on happiness or life satisfaction. A famous example is offered by Blanchflower and Oswald (2004). The authors estimate that a lasting marriage (compared to widowhood as a natural experiment) is worth \$100,000 a year. As far as we know, there are no applications in the equivalent income tradition yet.[34] It would not be difficult to derive equivalent incomes on the basis of a life satisfaction equation, however, and the marginal rate of substitution estimated by Blanchflower and Oswald shows that the willingness-to-pay for a good family life is likely to be considerable.

In these studies, the ultimate goal is to measure well-being as an aggregate over many dimensions. This has also been the perspective of Section 2.3. It is instructive to compare this perspective to that taken by the large body of literature that tries to calculate so-called equivalence scales. The basic question to be answered by this approach is the following: "How much income does a household with characteristics z need to reach the same level of well-being as a reference household?" where the latter is usually—but not always—taken to be a single. Therefore, the proclaimed ambition of this literature is also to compare the well-being of different households. The problem that researchers working in this field want to tackle is that income (and consumption) are usually reported at the level of the household and not at the level of the individual. Yet, it is obvious that living in a household involves returns to scale, including the consumption of household public goods. Think about housing or about the use of a car, for instance. It is natural to assume that a couple needs less than twice the income of a single to reach the same level of well-being. The challenge is then to try to correct reported incomes at the household level to take into account differences in household composition.

This is a very old problem, about which no consensus has been reached yet. As a matter of fact, despite the large academic literature on the topic, most practitioners are still using equivalence scales without a coherent theoretical foundation. A typical example is the so-called modified OECD scale used by Eurostat, in which the first adult counts for 1, the second adult and each subsequent person aged 14 and over counts for 0.5, and for each additional child under 14 one adds 0.3. The household income is then divided

[34] Household size plays an important role in the application of Fleurbaey and Gaulier (2009), but this application is at the country level—see Section 2.5.3.

by this scale to get the "equivalized income."[35] Alternatively, the OECD divides the household income by the square root of the household size. In both cases the reference household (for which the equivalized income equals the original income) is a single. The lack of consensus about the exact scale to be used has also stimulated the use of stochastic dominance approaches (Atkinson and Bourguignon, 1987; Bourguignon, 1989; Fleurbaey et al., 2003; Ooghe and Lambert, 2006). We will not summarize the large literature on equivalence scales here, but rather focus on the differences and similarities with the approaches to measuring well-being that are the topic of this chapter.

Using the cost function $C(u, p, z)$ to denote the minimum expenditure needed by a household with characteristics z to reach utility level u if prices are p, and denoting the reference household characteristics by $\bar{z}$, the equivalence scale is defined as

$$\mathrm{ES}(u, p, z) = \frac{C(u, p; z)}{C(u, p; \bar{z})}, \tag{2.10}$$

and the equivalized income as

$$y_E(u, p, z) \equiv \frac{y}{\mathrm{ES}(u, p, z)} = y\frac{C(u, p; \bar{z})}{C(u, p; z)} = C(u, p; \bar{z}).$$

By far most attention went to the derivation of equivalence scales on the basis of observed consumption behavior. Traditionally, the analysis of consumption behavior was based on the assumption of a "unitary" household, with preferences and optimization behavior defined at the level of the household. To go beyond the household level and compute individual well-being, it was then commonly assumed that all household members experience the same well-being level. It is clear that in this approach the calculation of equivalence scales requires interpersonal comparisons of well-being between households of different sizes. It is not easy, however, to give an intuitively attractive interpretation to well-being at the level of the household. More important, it is immediately obvious that consumption data do not yield sufficient information to allow for such interhousehold comparisons of well-being. More specifically, what we can (under some conditions) identify are different sets of indifference curves (one for each household type), but observed consumption does not give us any clue about how to link these indifference curves to utility levels. Stated more formally, the cost functions $C(u, p, z)$ and $C(\delta(u, z), p, z)$ will induce exactly the same consumption behavior—where the transformation $\delta(u, z)$ may depend on z.

Identification of the equivalence scales can only be achieved by introducing additional assumptions. The most famous of these is the so-called IB assumption, where IB stands for "independence of base" (Lewbel, 1989).[36] This assumption states that

[35] We use the term *equivalized income* to denote the income divided by an equivalence scale. This is to be distinguished from the equivalent income that was introduced in Section 2.3.

[36] The same assumption has been proposed by Blackorby and Donaldson (1993) under the name "equivalence-scale exactness." It has later been generalized by Donaldson and Pendakur (2003), but this generalization does not solve the basic issue described here.

the equivalence scale is independent of utility, i.e., $C(u, p; z) = C(u, p; \overline{z})\mathrm{EB}(p, z)$, where $\mathrm{EB}(p, z)$ refers to an equivalence scale that satisfies the IB assumption. This assumption implies a restriction on the cost functions and therefore leads to testable restrictions on the consumption behavior of different households. A crucial part of the identifying assumption is not testable, however, notably the assumption that all households with the same value for $y/EB(p, z)$ indeed reach the same level of utility.

How to interpret this approach in the light of our broader questions about well-being? First, the concept of well-being used is a restricted one. In fact, as argued in the short but influential paper by Pollak and Wales (1979), equivalence scales as derived from consumption behavior do not include the direct effects of z on utility—and unless one includes choice of household size in the analysis, choice behavior can never reveal any information about preferences with respect to household size. Pollak and Wales draw a distinction between *situation comparisons* and *welfare comparisons*. Situation comparisons are based on the conditional cost function, giving the minimum expenditures needed to reach a given utility level u, conditional on having characteristics z. Welfare comparisons, on the other hand, require the estimation of an unconditional cost function, giving the minimum expenditures needed to reach a given utility level u, taking into account the direct effect of the characteristics z on utility. In fact, Pollak and Wales are critical of the traditional approach and state that "conditional equivalence scales estimated from observed differences in the consumption patterns of families with different demographic profiles cannot be used to make welfare comparisons" (Pollak and Wales, 1979, p. 220). Unconditional utility (or cost) functions are the representation of preferences over bundles of life dimensions, with household characteristics z as elements of those bundles. Pollak and Wales therefore reject the relevance of traditional equivalence scales and advocate the use of the methods that have been described in the previous sections.

Many (or even most) authors studying equivalence scales take a less-negative position and argue that situation comparisons, despite their limitations, are meaningful on their own. They do not yield real welfare comparisons because they do not take into account the direct effect of family life (having a partner and children) on well-being. However, they do make sense in a resource-based approach, focusing on incomes and material consumption only. Some would even claim that inequality in material welfare, as measured by equivalized incomes, is more relevant for policy purposes than inequality in overall well-being, as it is not generally accepted that households should be compensated, e.g., for the fact of having children or not. The relevant question then becomes whether the IB-assumption that the equivalence scale is independent of utility is attractive from a normative point of view. This turns out to be a difficult question as the concepts of preferences and utility are difficult to interpret when applied at the level of the household.

As a matter of fact, this issue extends beyond the problem of correcting for household size. Similarly, one can calculate equivalence scales for other characteristics z. As an example, Jones and O'Donnell (1995) presented equivalence scales for disability, focusing on "the extra expenditure required by a household with a disabled person to achieve the same level of welfare as a reference household without any disabled individuals." In such a context of disability, the distinction between situation and welfare comparisons seems even more relevant, although (as noted by the authors) in this setting one can consider these extra expenditures as a lower bound on the welfare loss resulting from disability.

A major drawback of the traditional approach is its assumption that preferences and welfare can be defined at the level of the household. It is much more natural to see the household as consisting of individual members, each with their own individual preferences and deciding jointly about household consumption.[37] In this respect, an important recent breakthrough has been the move from the "unitary" to the "collective" model of household behavior (Apps and Rees, 1988; Chiappori, 1988). Some goods are purely private (food or clothing), others are public, and some may be mixed (a car can be used by all household members to make a trip jointly, but it can also be used by only one member of the household). Household resources are allocated to the consumption of each of the household members on the basis of a sharing rule. This rule reflects the relative power positions of the different household members. Finding the restrictions needed to identify the individual preferences of the household members and the sharing rule on the basis of observed consumption behavior in a setting with joint consumption and externalities is a very active and rapidly expanding field of research. This literature is discussed extensively in Chapter 16. Here we focus on the crucial relevance of this work for measuring individual well-being. Indeed, moving from the unitary to the collective model is an important advance in this regard.

A first approach to measuring individual well-being focuses on the sharing rule. If one reasons within a resource-based approach, the share of resources devoted to the consumption of individual *i* (as influenced by the distribution of power within the household) is an important indicator of his relative well-being level. Identification of the level of the sharing rule is not easy and requires additional restrictions, but Cherchye et al. (2013) show that upper and lower boundaries can be identified in a nonparametric setting. Applying their method to observations in the 1999–2009 Panel Study of Income Dynamics on childless couples, where both adult members participate in the labor market, leads to some interesting insights. As an example, while 11% of their restricted sample have incomes below the two-person poverty line; between 16% and 20% of individuals are below the individual poverty line. Using semiparametric restrictions, Dunbar et al. (2013) identified the resource shares of different household members, including children. Using data for Malawi, they found that the overall poverty rate calculated at the household level understates the incidence of child poverty.

The sharing rule-approach defines well-being in terms of income. For our purposes, another application of the collective approach is more relevant, however. Given that

[37] Early examples of this approach are Manser and Brown (1980) and McElroy and Horney (1981).

(under some assumptions) it is possible to identify individual preferences, one can formulate an answer to the question: "How much income would an individual living alone need to attain the same indifference curve over goods that the individual attains as a member of the household?" (see, e.g., Browning et al., 2013; Lewbel, 2003). There is an essential difference between this question and the one that was formulated earlier in the context of traditional equivalence scales. To answer this question, one just needs information about the indifference map of the individual, without having to label them. The resulting so-called individual indifference scales are closely related to the notion of equivalent income because they obviously are a form of money-metric utility that can be calculated on the basis of ordinal preference information only. Indifference scales will depend on these preferences, on the "consumption technology" used by the household (in terms of private, public, and mixed goods), and on the sharing rule, i.e., the distribution of power within the household. Applications of this approach have focused, among others, on the adequate compensation in case of wrongful death (Lewbel, 2003) and on poverty among the elderly (Cherchye et al., 2012).

Although the introduction of the collective model constitutes an important step forward, it does not bridge the gap between welfare and situation comparisons. Indifference scales do not capture the direct utility effects of partnership and children and remain therefore situated within a resource-based approach. They therefore do not yield a complete measure of well-being taking all relevant life dimensions into account. Whether one considers this to be a problem or not depends on whether one thinks that resource-based (situation) comparisons are relevant from a policy point of view.

Until now, we have discussed the approach to equivalence scales that focuses on observed consumption behavior. Because the focus is on identifying individual preferences, this approach is close to the intuitions underlying the equivalent income approach. Let us now see how the two other approaches to well-being measurement have been applied to tackle the equivalence scales problem.

There are almost no applications within the capabilities framework. Lelli (2005) calculated the equivalence scale of a household with characteristics z as the income needed to reach the same level of functioning (in her case housing) as the reference household. Her application thus remains limited to one functioning—and the resource-based perspective underlying this analysis goes in fact against the basic inspiration of the capability approach.

The subjective (or satisfaction) approach has been used more extensively for the construction of equivalence scales. The pioneering work in this field has been done by Van Praag (1971) and Kapteyn and Van Praag (1976). Originally, these authors assumed that there was a cardinal utility function of income $U(y; z)$, where z represents—as before—all relevant nonincome variables. They assumed (on the basis of a theoretical reasoning) that this utility function takes the form of a lognormal distribution function $U(y; \mu(y^A, z), \sigma(y^A, z))$, with the mean and the standard deviation dependent on z and on the actual income y^A of the household. The parameters of that function were estimated on the basis of the answers obtained from what was called the "Income Evaluation Question" (Van Praag, 1971). This question goes as follows:

> *Please try to indicate what you consider to be an appropriate amount for each of the following cases. Under my (our) conditions I would call a net household income per week/month/year of:*
> *about__________________very bad*
> *about__________________bad*
> *about__________________insufficient*
> *about__________________sufficient*
> *about__________________good*
> *about__________________very good.*

Giving specific values to the labels allowed them to estimate the utility function and hence to derive equivalence scales as in expression (10). This original Van Praag approach has never become very popular, partly because of the strong assumptions of cardinality and lognormality.

In later work (e.g., Van Praag and van der Sar, 1988), the cardinality assumption was dropped and the only assumption that was retained was that the different labels (from "very bad" to "very good") correspond to the same utility values for all individuals (i.e., that they were interpersonally comparable). Denoting the answers given by individual i for label k by c_{ik}, Van Praag and van der Sar (1988) then specified and estimated a (loglinear) function $c_{ik}(y_i, z_i)$:

$$\ln c_{ik} = \beta_{0k} + \beta_{1k} \ln z_i + \beta_{2k} \ln y_i + \epsilon_{ik},$$

where z_i is the size of individual i's household and ϵ_{ik} is an error term. The coefficients β_{2k} turn out to be highly significant. Respondents with a higher income evaluate the income needed to reach a given utility level as significantly higher than respondents with a lower income. Van Praag talks about a "preference drift" effect, and there is clear echo of the phenomenon of adaptation that has been discussed before. Taking this preference drift into account, one can derive that the "true" cost level needed to reach utility level k is found where $c_{ik} = y_i$, i.e.,

$$\hat{c}_k(z) = \exp[(\beta_{0k} + \beta_{1k} \ln z_i)/(1 - \beta_{2k})].$$

The equivalence scale at level k can then be calculated as $\hat{c}_k(z)/\hat{c}_k(\overline{z})$, where $\overline{z}$ again denotes the reference household. In their sample of eight European countries and the United States, the equivalence scales are reasonably similar at the different k levels, which gives some support for the IB assumption used in the consumption approach.

Other authors (e.g., Koulovatianos et al., 2005) have implemented a similar method with different formulations of the subjective question. A few papers have combined consumption data and subjective questions (de Ree et al., 2013; Kapteyn, 1994). This is a promising approach because it allows identification of preference parameters on the basis of the subjective information. As an example, de Ree et al. (2013) rejected the IB assumption (and its generalizations) for a sample of Indonesian households.

The question arises of whether the subjective method yields welfare or situation comparisons. Surely, the analyses on the basis of overall life satisfaction that we discussed in the beginning of this section yield welfare comparisons. This is much less clear for the subjective questions used in the literature on equivalence scales, however. Do individuals responding to the income evaluation question take into account the direct effect of household size on well-being? They probably do not, but it is not fully clear for the income evaluation question given earlier. Koulovatianos et al. (2005) confront their respondents with hypothetical household situations and then ask: "Given that someone has an extra child, how much would they need to reach the same level of well-being?" They argue that this yields conditional scales. The subjective information used by de Ree et al. (2013) is even more related to adequacy of resources. The subjective approach to equivalence scales estimation therefore also yields only situation comparisons—and deliberately so.

We can conclude that most of the studies within the equivalence scale approach do *not* aim at comparisons of well-being, taking into account at the same time the effects of income and of the quality of family life. On the contrary, they aim at needs-corrected values of income, i.e., at conditional comparisons of well-being. They are therefore rooted in a resourcist view on well-being. However, the methods that have been used to calculate equivalence scales are similar to the methods that we described in Section 2.3. The capability approach has hardly been used in this context. Equivalence scales derived from consumption behavior are based on preferences. Although the traditional literature based on the unitary model requires arbitrary assumptions about interpersonal comparability, the recent work with the collective model derives indifference scales using only ordinal preference information. The intuition underlying this approach is closely related to that behind the concept of equivalent income. Subjective evaluations have also been used, and in some of the work there is explicit consideration of the adaptation phenomenon. We will see in the next section that the issue of welfare versus situation comparisons is also relevant to interpret the literature on publicly provided services and benefits.

2.5.2 Publicly Provided Services and Benefits

Countries differ in the extent to which services are provided publicly rather than through the private market, for instance. Comparing the income distribution of two countries, one where health services are primarily covered by private out-of-pocket payments and another where such services are provided free of charge, may result in misleading conclusions on which country is preferable from a social welfare point of view. In addition, publicly provided services and benefits may have an impact on the inequality within a country.[38]

[38] The publicly provided services that are typically covered in the applied literature include educational benefits, health care, social housing, food stamps, and child care. On average across OECD countries, the first two listed services are estimated to add up to about 13% of GDP, ranging from 8% in Turkey up to 20% in Denmark and Sweden (Verbist et al., 2012).

As a starting point, it is fruitful to recall the distinction between functionings and resources. We argued that what matters to define well-being are the functionings of a person, i.e., that person's "beings" and "doings." Resources, on the other hand, can be used to achieve certain functionings. These two concepts are different, as individuals may differ in how they convert resources into functionings. An analysis of well-being inequality using a broad set of functionings as the relevant space of well-being includes automatically the publicly provided services, insofar as they contribute to the functionings of the individuals. This approach seems natural in view of the discussions of this chapter. Yet, to the best of our knowledge, examples of this direct method to include publicly provided services into distributional analysis are scarce. Instead, a resource-based approach is standard practice in the literature. In this method, disposable income is extended with a monetary valuation of the publicly provided services. The resulting measure of extended income uses a monetary valuation of the external resources that individuals have at their disposal to obtain functionings and to reach well-being. The inspiration of this approach is therefore closely related to that of equivalence scales as described in the previous section.

In this section, we will first shortly survey the popular approach to extend disposable income with information on publicly provided services. Then we will discuss the issue of the valuation of public services and the adjustment for needs in view of the normative issues discussed in this chapter.[39]

2.5.2.1 The Extended Income Approach

The extended income approach consists of three steps. In a first step, one *selects* the government services to be included. Then, second, these services are *valued* at their production cost for the government. Finally, the value of the service is *allocated* to the beneficiaries based on their actual consumption or the insurance value, depending on the benefit at hand.[40] The obtained value of the monetary value of the publicly provided service is added to the disposable income of the household to obtain its extended income. The distribution of extended income is then analyzed with standard inequality measures.

Applying an extended income approach, the OECD flagship report "Growing Unequal?" (OECD, 2008; Chapter 8) has obtained the following findings. First, the inclusion of publicly provided services reduces income inequality within countries

[39] More extensive surveys can be found in Smeeding (1982), Marical et al. (2008), and Verbist et al. (2012). In this section we do not discuss intergenerational equity concerns involving publicly provided services (see Bourguignon and Rogers, 2007, however).

[40] The insurance value method is commonly used in the case of health care services, where the allocation is based on the average spending on the relevant age-sex group irrespective of the actual use that was made of the service. This method interprets health care as an insurance benefit received by all covered individuals. The value of the insurance benefit approximates an actuarially fair insurance premium assuming that all individuals with the same age-sex characteristics are paying the same premium.

because of their predominantly uniform character. Yet, this reduction is typically lower than the inequality reduction obtained by tax and cash benefits. Second, the differences in income inequality between countries are reduced as well, but the ranking of the countries with respect to extended income inequality remains similar to the ranking according to income inequality (affirming earlier findings of Smeeding et al., 1993).

The extended income approach has the advantage of being implementable for many countries. It requires income data that are readily available in standard household surveys and some additional macroestimates of the production cost of public services. Yet, in view of the normative issues raised in this chapter, the extended income approach appears to take an overly pragmatic view on the valuation of the contribution of publicly provided services to well-being.

2.5.2.2 Valuing Publicly Provided Services and Respect for Preferences

How a publicly provided service should be valued depends arguably on the purpose of the valuation exercise. Whereas valuing the benefit by means of its production cost may give a good estimate of its budgetary cost, this valuation method seems less appropriate for the purpose discussed in this chapter (i.e., an analysis of the distribution of individual well-being).[41]

An example illustrates why this is the case. Imagine that the value of publicly provided education benefits is determined by their production cost. An increase in the wages of teachers increases the production cost, but it seems counterintuitive to say that the value of the service for the recipients has increased *because* the production cost has increased. Indeed, this valuation method neglects the efficiency of the production process and potential quality differences between equally expensive services. A valuation by means of the production cost can therefore best be seen as an approximation when no other information is available. Alternative valuation methods have been proposed that are more closely related to the preferences of the population.[42]

A first alternative would be to use the *market value* of the public services rather than the production cost. Under some circumstances market prices may indeed give an indication of willingness-to-pay. Of course, this method can only be applied to the publicly

[41] Note that in an approach that values publicly provided services by the production costs, the measured productivity growth is always equal to zero (because productivity equals the ratio of output to input). The so-called Atkinson Review (see Atkinson, 2005; Pont, 2008) provides some general principles for an improved measurement of aggregate government output that captures productivity and quality improvements.

[42] See Smeeding (1982) for a more extensive survey. In the 1980s, the market valuation and cash equivalent valuation have played a central role in the computation by the U.S. Census Bureau of an "experimental" poverty measure for the United States, including the value of publicly provided services (Fisher, 1997). More recently, the production cost approach seems to have become the gold standard in the applied literature, as can be witnessed from the survey by Marical et al. (2008).

provided services for which a private market exists. For food stamps and social housing, for instance, the market value can either be inferred directly or obtained by a hedonic regression. However, the prices of privately provided services do not always reflect the valuation of the recipients either. Stiglitz et al. (2009, p. 99) give an example in the market for privately provided medical services where informational asymmetries disconnect market prices from marginal valuations and preferences.

A second alternative valuation method relies directly on the preferences of the recipients and measures the value of the publicly provided service by its so-called cash-equivalent. That is the amount of cash needed to induce an individual to forgo a particular publicly provided service (see, Smeeding, 1977, for instance). Insofar as the individual's own preferences (her own willingness-to-pay) are used to compute these cash equivalents, this method respects the same-preference principle. Figure 2.6 illustrates the cash-equivalent valuation method graphically in the income-health space. Consider two individuals, Alexandra and Benny, who are equally rich (their income equals OA on the graph). Alexandra is in better health than Benny (their health is respectively OG and OE). Both individuals receive publicly provided health services, without which Alexandra's health would be OF, whereas Benny's health would be only OD. It is clear from the figure that the publicly provided health services generate a larger increase in the health of Benny than they do for Alexandra. However, Benny cares relatively less about his health than Alexandra does (as Benny's indifference curve is "steeper"). Benny's cash equivalent for the health service is AB, as Benny is willing to forgo the publicly provided health service for an additional income of AB. Alexandra's cash equivalent, on the other hand, equals AC. Even if the health service generates a smaller health increase for Alexandra, her cash equivalent is larger, as she cares more about health than Benny does.

Estimating a cash equivalent requires additional information compared to the production cost approach. As illustrated by the preceding example, one needs to know the preferences of Alexandra and Benny because the magnitude of their cash equivalent is determined by the shape of their indifference map. To obtain the necessary information on preferences, the methods surveyed in Section 2.3 can be used, i.e., revealed preferences, stated preferences, and satisfaction data. The revealed preferences method seems to be favored in the applied literature. Typically, the preferences are derived from consumption behavior by means of an estimated system of demand equations (Slesnick, 1996; Smolensky et al., 1977). Life satisfaction data can also be used to estimate willingness-to-pay for publicly provided services. Levinson (2012), for instance, estimated the willingness-to-pay for air quality and computes the compensating variation for air pollution.

Typically the cash-equivalent method is not formulated in the space of functionings (or life dimensions), and it does not yield an overall measure of well-being that is based on a coherent ethical reasoning. We will come back to this issue in the following subsection. However, because it focuses on the individual willingness-to-pay, its inspiration remains

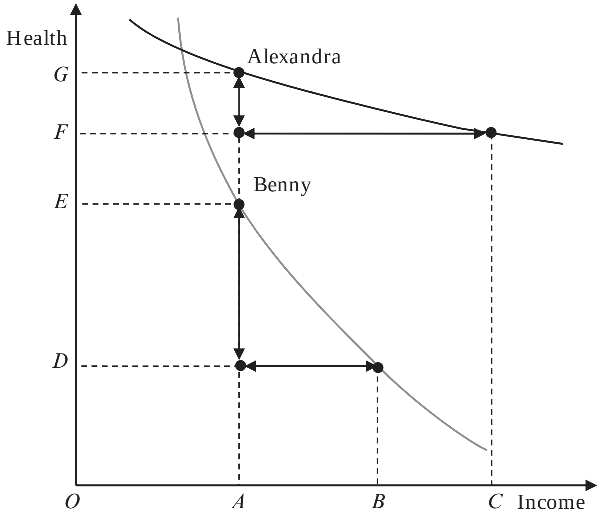


Figure 2.6 Cash equivalent income.

similar to that of the equivalent income approach to well-being as defined in Section 2.3.3. Other approaches reject completely the idea that individual preferences and willingness-to-pay provide the best guidelines to value publicly provided services.

Governments may provide these in-kind services exactly because they are inspired by paternalistic motives or by a concern about consumption externalities (see Currie and Gahvari, 2008, and the references therein for a detailed discussion). The paternalistic motivations reflect Musgrave's (1959) idea that some goods are merit goods, which leads to an immediate conflict with the idea of respecting individual preferences. A paternalistic government values the publicly provided services according to an objective valuation function, which requires arguably a perfectionist or objective theory of well-being. As suggested before, the gap between these two approaches may be bridged to some extent by introducing a distinction between informed and uninformed preferences.

2.5.2.3 Adjusting for Needs and Individual Responsibility

Because extended income focuses on the willingness-to-pay for the services, it does not at all take into account the functioning levels that the individuals would obtain in absence of any publicly provided service. This can be seen using Figure 2.7, depicting the functionings and indifference curves of Alexandra and Charlotte in the income-health space. Alexandra and Charlotte have the same income (OA) and obtain the same increase in their health from the publicly provided health services ($ED=FG$). Moreover they have the same preferences, so their cash equivalents are equal to AC.[43] Hence, the extended

[43] The argument does not depend on the choice for a particular valuation method for the publicly provided service.

income of both individuals is equal (OC). Consequently, when extended income would be used as a measure of well-being, both individuals would be considered as equally well off. Yet, no account is taken of the fact that the health levels that the individuals would obtain in absence of any publicly provided health service may be very different (Alexandra is in much better health than Charlotte, $OF > OD$). Radner (1997) illustrated a similar issue by showing how the well-being of elderly (Charlotte in Figure 2.7) would be overestimated as the value of publicly provided services is included in their extended income without taking account of their needs. This observation led Paulus et al. (2010, p. 263) to doubt whether results derived using the extended income approach can have a straightforward welfare interpretation.

This discussion echoes the distinction that was introduced in the previous section between welfare and situation comparisons. As a matter of fact, in the recent applied literature, the solution has been sought in adjusting the equivalence scales of the recipients for differences in needs.[44] Paulus et al. (2010) adopted a "fixed cost" approach, in which the needs of a recipient are assumed to be equal to a specific fixed monetary amount. In particular, it is assumed that the per capita amounts spent for age-specific population groups on public services accurately depict the corresponding needs of these groups. Under this assumption an equivalence scale for extended income can be inferred for each household. The authors then performed a sensitivity analysis of the inequality reducing

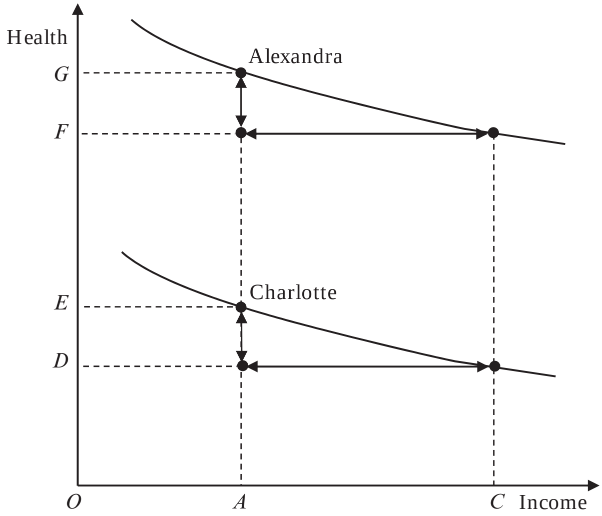


Figure 2.7 Cash equivalent income and needs.

[44] It should be noted that Smeeding et al. (1993) applied different equivalence scales to disposable income and the value of the publicly provided services. A standard equivalence scale based on household size and number of children is applied to disposable income. The publicly provided services received by the household, on the other hand, are distributed on a per capita basis over all its members.

effect of public services from changing the amount of received publicly provided services to the European average for the age-specific groups.

Aaberge et al. (2010) derived needs-adjusted equivalence scales consistent with their preferred allocation method of the production costs across target groups (i.e., a model of spending behavior of local governments). The equivalence scale for noncash income is obtained from the estimates of the relative needs of different target groups that are derived from the minimum expenditures identified in the spending model. Using data from Norway, they find that including publicly provided services reduces income inequality considerably, but that adjusting for needs offsets about half of the inequality reduction.

The method hinges on two strong assumptions on the interpretation of the spending model for the local governments (Aaberge et al., 2010, p. 552). First, the estimated minimum expenditures are to be interpreted as originating from an implicit consensus among local governments about how much spending the different target groups need minimally. Second, the functional form of the individual well-being measure derived from public services is assumed to coincide with the functional form used by local governments to decide the spending on public services. A priori, it seems hard to square such a (heroic) assumption with the idea of respecting individual preferences.

Both approaches to compute a needs-adjusted extended income rely on a two-step procedure. In a first step, the extended income is computed by adding a monetary value of the publicly provided services to the disposable income of the individual. Then, in a second step, these extended incomes are adjusted for differences in individual needs by means of an equivalence scale. A natural alternative would be to measure well-being directly in the desired space, i.e., functionings or capabilities themselves. For that purpose, a well-being measure should be developed along the lines described in Section 2.3 of this chapter. The relationship between these broader measures and the resource-based measures that are used now deserves a deeper exploration.

2.5.3 International Comparisons

The international comparison of living standards is fraught with many difficulties (see also Chapter 11). Because the focus in this literature has often been on comparing real incomes, we will first discuss the difficulties related to differences in prices for market commodities, and we will show how they are linked to the normative issues discussed in the previous sections. We will then show the relevance of introducing nonmarket dimensions into the evaluation of living standards.

2.5.3.1 PPP Indexes

International comparisons of living standards involve the search for price deflators that make it possible to compute comparable real incomes.[45] Pragmatic convenience

[45] Good introductions to this field are offered by Neary (2004), Deaton (2010), and Deaton and Heston (2010).

motivates approaches in which indexes are computed directly from prices and quantities, without depending on an estimation of consumer preferences. The theory of index numbers initiated by Fisher (1922) and developed by Diewert (1976, 1992a,b) is an important source of inspiration for such indexes. Pragmatic convenience also encourages seeking formulae that make the comparison of two countries independent of data from third countries.

Consistency is akin to transitivity in the comparison of real incomes, but a cardinal form of transitivity (involving orders of magnitude) appears desirable, not just an ordinal form. For instance, if Q_{ij} is a quantity index that compares real income in country i to country j, consistency is achieved when the chain relation $Q_{ij} = Q_{ik}Q_{kj}$ holds for every third country k. A popular way of achieving consistency computes real income as the value of quantities consumed at reference prices $\bar{p}$, so that $Q_{ij} = \bar{p}q_i/\bar{p}q_j$, where q_i is the vector of total consumption in country i.

All approaches have a connection with consumer preferences, but the conditions required are more or less restrictive, and the connection therefore more or less loose. There seems to be near consensus, in the PPP literature, that "in so far as data on real income have any meaning, it is that they provide an answer to the question: 'How well off would the same reference consumer be in different countries?'" (Neary, 2004, p. 1425). In other words, even if heterogeneous preferences may be the fact of the matter, there is no real attempt to formulate indexes that reflect this diversity of preferences. The main implicit underlying argument seems to be that ordinal preferences do not allow for interpersonal comparisons, unless arbitrary assumptions are made. In particular, money-metric utilities are not considered a possible option, although very similar notions are sometimes used, as explained later. This observation again suggests that the focus is not on welfare comparisons, but on situation comparisons.

Let us first briefly describe the two most popular PPP methodologies. The Eltetö–Köves–Szulc (EKS) quantity indices, used by the OECD and Eurostat are multilateral extensions of the Fisher index:

$$Q_{ij}^{\mathrm{EKS}} = \prod_{k=1}^{n}\left(Q_{ik}^{F}Q_{kj}^{F}\right)^{1/n},$$
$$Q_{ik}^{\mathrm{F}} = \sqrt{\frac{p_i q_i}{p_i q_k}\frac{p_k q_i}{p_k q_k}}.$$

They satisfy consistency in the form of the chain relation, but depend on third country data. They do not require estimation of preferences, and the link to consumer preferences is usually made by referring to Diewert's (1976) argument that the Fisher quantity index Q_{ik}^{F} is equal to the exact index $e(p_k, u(q_i))/e(p_k, u(q_k))$ of a flexible expenditure function $e(p, u)$, i.e., a function that approximates any twice differentiable expenditure function to the second order. Along a similar vein, Neary (2004) proved that Q_{ij}^{EKS} is equal to the ratio of utilities when utility is quadratic $u = \sqrt{q'Aq}$, for some suitably chosen symmetric matrix A. Again, a quadratic utility is a flexible form. These approximation results, however, are compatible with the index being sometimes wrong in the first order even for small changes—as must happen with any index that ignores preferences.[46]

Another popular approach, used by the U.N. International Comparison Project and the Penn World Table (PWT), relies on the Geary–Khamis (GK) indices, which compute PPP expenditures as the value of consumption at reference prices, $\bar{p}q_i$, and the reference prices are derived from the system

$$\bar{p}_k = \frac{\sum_i s_{ik}\bar{p}q_i}{\sum_i q_{ik}}, \tag{2.11}$$

where s_{ik} is the budget share of commodity k in country i. If one defines $\bar{s}_{ik} = \bar{p}_k q_{ik}/\bar{p}q_i$, one sees that the GK system can be written as

$$\sum_i \bar{s}_{ik}\bar{p}q_i = \sum_i s_{ik}\bar{p}q_i.$$

This approach obviously satisfies consistency. It depends on third country data, but only in the computation of $\bar{p}$. The link with consumer preferences is tenuous because $\bar{p}q_i$ provides a good index only for Leontief preferences (which would imply that all countries' consumption vectors should be proportional to one another). Neary (2004) then proposed, as a variant, to estimate world consumer preferences and substitute compensated demands q_i^* to actual quantities q_i for the computation of reference prices in system (11). Taking $\bar{p}q_i^* = e(\bar{p}, u(q_i))$ as the real income values is then truly faithful to the estimated preferences—but not necessarily to the population's actual preferences if they are heterogeneous, as noted in van Veelen and van der Weide (2008).

Interestingly, nothing in expression (11) as modified by Neary requires identical preferences, so that one could apply Neary's methodology to a population with country-specific estimated preferences, in which case the real incomes $\bar{p}q_i^* = e_i(\bar{p}, u_i(q_i))$ would be money-metric utilities at the country level. This idea is not considered in van Veelen and van der Weide (2008) or in the reply by Crawford and Neary (2008). As mentioned earlier, Fleurbaey and Blanchet (2013) proposed to take other reference prices for the computation of money-metric utilities, namely, prices $\bar{p}$ that maximize $\sum_i e_i(\bar{p}, u_i(q_i))/\sum_i \bar{p}q_i$. This minimizes the aggregate Gershenkron effect[47] and renders $\sum_i q_i^*$ proportional to $\sum_i q_i$.

van Veelen (2002) proved an impossibility theorem that is similar to the incompatibility between the personal-preference principle and the dominance principle discussed earlier in this chapter. This theorem says that there is no measure of real income (based on

[46] See Fleurbaey and Blanchet (2013, p. 95) for more details.

[47] The Gershenkron effect is the observation that the more p_i differs from $\bar{p}$ and q_i differs from q_i^* the more $\bar{p}q_i$ overestimates $\bar{p}q_i^* = e_i(\bar{p}, u_i(q_i))$.

prices and quantities in all countries) that is continuous, is not independent of prices, satisfies dominance ($q_i > q_j$ implies that real income is greater in i), and such that pairwise comparisons are independent of third countries.[48] The EKS and GK methods satisfy all conditions except the last one. A money-metric approach that estimates preferences on the same data may, in addition, fail to satisfy dominance in the case of heterogeneous estimated preferences. We know that this is a necessary consequence of respecting heterogeneous preferences.[49]

In a recent paper, Almas (2012) considers exploiting preference data by estimating budget coefficients with household surveys, but retains the assumption of identical preferences. Instead of estimating a complete system of demand functions to compute expenditure functions and money-metric utilities, she focused on food and assumed that the equation of food share, conditional on demographic characteristics, is the same everywhere. Estimating it with the PPP price indexes from the PWT, she included country dummies and assimilated these dummies to a bias in the PPP indices. This method relies on the assumption that preferences for food versus other goods are identical all over the world, and it is not indicative of welfare because incomes deflated with the corrected indices are *not* money-metric utilities for the AIDS model that is estimated.[50]

Deaton (2010) and Deaton and Heston (2010) studied the difficulties created by the fact that different countries in fact consume different lists of commodities, with great differences between countries with very unequal standards of living. The worst configuration would, of course, be the case in which every country consumes its own specific list, that has no intersection with the list of other countries. In this case, it is hard to imagine how to perform comparisons on the basis of observed market demand data. But even when all pairs of countries have a nonempty intersection of lists, the imperfect overlap creates difficulties. Practical methods that single out identical but often nonrepresentative goods appear unsatisfactory (a popular example, however, is the so-called Big Mac Index, published yearly by the *Economist*). Using proximate countries to compute chained indexes may lead to compounding errors as one compares distant countries. Deaton (2010) suggested that nondemand data, such as well-being questionnaires, may provide useful additional information for comparisons across countries. On the theory side, Fleurbaey and Tadenuma (2007) showed that imperfect overlap of commodity lists generates Arrow-like impossibility theorems, even if one only relies on the weak independence axiom stipulating that the evaluation of two allocations should only depend on

[48] See Quiggin and van Veelen (2007) for a further analysis of similar ideas.

[49] Note that if preferences were known on the basis of other data, then a money-metric approach with a fixed reference price would satisfy the last condition but would fail the price-dependence condition, as quantities would provide all needed information about welfare.

[50] In fact she does not estimate the AIDS model, but only the equation of food share in which the deflator of income is the PWT PPP index rather than the AIDS deflator.

preferences over the commodities that appear in either allocation. As a way out, they suggested focusing on lists of functionings that have a common set of core components, which is not very different from Deaton's suggestion to go beyond market data. Of course, these suggestions immediately bring us the more general topic of introducing nonmarket dimensions.

2.5.3.2 Nonmarket Dimensions

The recognition that living standards incorporate public goods of many sorts (e.g., the environment), as well as nonmarketed goods and "functionings" (e.g., health), has played an important role in the motivation to go "beyond GDP" not just for the evaluation of growth and public policy in a given country, but also in international comparisons. In fact, all three approaches that were reviewed in Section 2.3 have been applied in empirical work on intercountry comparisons.

The simplest approach consists in aggregating indices of the different dimensions of life into a single composite indicator. This approach follows the objective interpretation of the *capability approach*. The most popular example is the HDI that aggregates three indices (which are normalized between 0 and 1 from the range of achievements by the various countries of the world): national income, life expectancy, and education. Although the initial version of the index made a linear aggregation (UNDP, 1990) and therefore implied perfect substitutability between the dimensions, the geometric mean has recently been adopted to reflect the greater importance of a dimension when its level is low compared to the others (UNDP, 2010). In a variant of the new index, the average indices per domain can be adjusted for inequality, so as to make each index a geometric mean of individual achievements. In this fashion, the global index can then also be written as the geometric mean of individual Cobb–Douglas indexes, due to the following identity (where I_i, L_i, E_i denote income, life expectancy, and education for individual i):

$$\prod_i \sqrt[3]{I_i L_i E_i} \equiv \sqrt[3]{\left(\prod_i I_i\right)\left(\prod_i L_i\right)\left(\prod_i E_i\right)}.$$

This variant alleviates the criticism raised against specific composite well-being indicators, that they fail to take the correlations between the dimensions or cumulative deprivation into account as they start from dimension-by-dimension summary statistics (see Section 2.4). In the preceding formula, the same elasticity of substitution is applied in the aggregation across dimensions and across individuals, so that the sequencing of both aggregations does not matter. This makes the index impervious to correlations between dimensions. Moreover, the present version of the HDI is clearly an objective index which—according to some—implies troubling trade-offs between the dimensions (see Ravallion, 2012, for instance).

There are many composite indicators that mimic the HDI methodology.[51] Some focus on social issues, whereas others focus on sustainability issues. The key difficulty for such indices is the choice of the weighting system for the various dimensions. It is quite common to perform sensitivity analysis to ascertain the robustness of conclusions to the weights (Decancq and Ooghe, 2010; Foster et al., 2013), which boils down to a dominance analysis. Another approach is to give up the aggregate index altogether and immediately apply multidimensional inequality indices to the same data (Decancq et al., 2009). Of course, as we saw in Section 2.4, none of these approaches allows to respect international preference heterogeneity.

The *happiness approach* has also been used for international comparisons, although much of the literature on cross-country data has focused on the link between happiness and income (Deaton, 2008; Diener et al., 2010; Stevenson and Wolfers, 2008). The great variations in average satisfaction with life at any given level of income may reflect differences in nonmarket dimensions of life, but also cultural variations. Helliwell et al. (2010) studied a large sample of countries and derived two conclusions. First, nonmarket dimensions play a large role in econometric regressions of life satisfaction. Such dimensions include having a partner, being able to count on friends, having freedom to choose, not perceiving corruption around oneself, having been generous, and practicing religion. These dimensions play a role at the individual level, but for some of them the national average also plays a role (including healthy life expectancy, which is not observed at the individual level). The second conclusion is that once one incorporates these social dimensions in the analysis, in a single equation of satisfaction with the same coefficients for all countries, the difference between predicted and actual values for the average level of satisfaction per country is small for most countries and has no systematic pattern, with one exception: the Latin-American countries in general have a higher well-being than predicted.

However, the results of country and regional equations also show that the coefficients of income and social dimensions vary substantially, revealing that the association between life satisfaction and the various dimensions of life is heterogeneous over the world. One can suspect that interpersonal heterogeneity may be even more important. If the satisfaction equations can be interpreted as giving some evidence on population preferences, this raises the interesting issue of comparing the situations of populations with different preferences—an issue that has been central in this chapter. Helliwell et al. (2010) proposed to compute income equivalent variations via the ratios of coefficients of social dimensions over the income coefficient. This method is one of those that have been introduced in Section 2.3 to estimate preferences necessary to calculate equivalent incomes.

[51] Surveys can be found in Gadrey and Jany-Catrice (2006), Stiglitz et al. (2009), and Fleurbaey and Blanchet (2013).

The method of *income equivalent variations* has been used by Becker et al. (2005) to estimate the income growth that would have been equivalent to the observed increase in life expectancy for various countries of the world. Their main finding is that the large increase in life expectancy in developing countries, once converted into a monetary equivalent, produces a much rosier picture of world inequalities than standard income measures. They assumed homogeneous preferences in the world, and their estimation of preferences relied on U.S. data on revealed preferences about job risks.

A combination of equivalent variations and compensating variations has been used by Jones and Klenow (2010), with a preference relation similar to that used in Becker et al. (2005), but extending the list of nonincome dimensions to include leisure time and inequalities. Letting I and Q denote income and quality of life (life expectancy, leisure, inequalities) and V a utility function representing the preference ordering (assumed to be common across countries), the equivalent variation approach solves the following equation, for each country i:

$$V(I_i, Q_i) = V\left(\lambda_i^{\mathrm{EV}} I_{\mathrm{USA}}, Q_{\mathrm{USA}}\right),$$

while the compensating variation approach solves the equation

$$V(I_{\mathrm{USA}}, Q_{\mathrm{USA}}) = V\left(I_i/\lambda_i^{CV}, Q_i\right).$$

They then propose to take $\sqrt{\lambda_i^{\mathrm{CV}}\lambda_i^{\mathrm{EV}}}$ as the index for comparisons across countries. A difficulty with the compensating variation approach, as they implement it, is that one may have $V(I_i, Q_i) > V(I_j, Q_j)$ but $\lambda_i^{\mathrm{CV}} < \lambda_j^{\mathrm{CV}}$. This problem is avoided with their equivalent variation approach because $\lambda_i^{\mathrm{EV}} I_{\mathrm{USA}}$ is a money-metric index based on quality of life in the United States as the reference.

Compensating and equivalent variation approaches are in general problematic when they make references vary with the object of comparison. Money-metric indexes avoid that difficulty by taking a fixed reference. Fleurbaey and Gaulier (2009) adopted a money-metric approach for international comparisons of OECD countries, with nonincome dimensions including leisure, life expectancy, unemployment risk, household composition, and income inequalities. They allowed for heterogeneous preferences for leisure only and relied on the Becker et al. (2005) preference ordering otherwise. Although the approach is in theory compatible with heterogeneous preferences at the individual level and the computation of a distribution of equivalent incomes within each country, they only focused on average levels for each country. Decancq and Schokkaert (2013) calculated individual equivalent incomes on the basis of the life satisfaction data from the European Social Survey with as nonincome dimensions health, employment status, quality of social interactions, and personal safety. They introduced these equivalent incomes into a concave social welfare function and compared the social welfare of 18 European countries for the years 2008 and 2010, taking into account the distribution of individual well-being. The ranking of the different countries in terms of equivalent incomes is different from the ranking of countries in terms of income. A striking example is the dramatic fall in the well-being of Greece and Spain as a result of the economic crisis. Bargain et al. (2013) studied heterogeneous preferences over consumption and leisure in various European countries and the United States and computed several money-metric indexes for the analysis of welfare level and inequalities. In their analysis also, preference heterogeneity plays an important role in the welfare rankings.

2.6. CONCLUSION

Egalitarian thinkers are usually concerned about the distribution of well-being. Individual well-being depends not only on income but also on other dimensions of life, such as health, the quality of social relations and of the environment, employment, and job satisfaction. In this chapter we have surveyed the economic literature on how to construct such overall measures of well-being. We distinguished three approaches: the capability (and functionings) approach, the use of subjective life satisfaction measures, and the calculation of equivalent incomes. We argued that the choice of measure ultimately is a normative issue, and we discussed the normative assumptions underlying the measurement of individual well-being, focusing on two issues: the degree to which individual preferences are respected and where in each approach the boundaries of individual responsibility are drawn. The three approaches take a different stance on these issues. We also compared the measurement of inequality in well-being with the use of multidimensional inequality measures. The latter only fit in a perfectionist perspective, completely neglecting interpersonal preference differences.

In most of the applied work on inequality measurement the ambition is more limited. One keeps focusing on resource-based measures, which are then extended to include other considerations: household size and composition (and other needs) in the literature on equivalence scales, the value of publicly provided goods and services, or differences in prices in the contect of international PPP comparisons. In each of these cases one usually does not aim at constructing an overall measure of well-being. One neglects (respectively) the direct effects of family relations on well-being, the attainment of functionings as a result of the public provision of goods and services, and the effect of international preference heterogeneity. In all these three domains one focuses on situation rather than welfare comparisons. However, the most common approaches are not really satisfactory, even from this more limited perspective, and the proposals to improve on these existing measures (the construction of indifference scales, the use of subjective satisfaction information, the introduction of willingness-to-pay and differences in needs in the context of public service provision, the introduction of preference differences in international comparisons) move the approaches in the direction of the construction of more global well-being measures and use methods that have also been explored and developed for the latter purpose. In fact, in some cases, the informational requirements become similar. Analyzing the exact relationship between "extended (or corrected) incomes" and overall measures of well-being is a fruitful area for further research.

ACKNOWLEDGMENT

We are grateful to the editors and John Roemer, Tim Smeeding, Tom Trimpeneers, Aki Tsuchiya, Gerlinde Verbist, and Frederic Vermeulen for helpful comments and suggestions on an earlier version.

REFERENCES

Aaberge, R., Bhuller, M., Langørgen, A., Mogstadt, M., 2010. The distributional impact of public services when needs differ. J. Pub. Econ. 94, 549–562.
Alesina, A., Di Tella, R., MacCulloch, R., 2004. Inequality and happiness: are Europeans and Americans different? J. Pub. Econ. 88, 2009–2042.
Alkire, S., 2002. Dimensions of human development. World Dev. 30, 181–205.
Almas, I., 2012. International income inequality: measuring PPP bias by estimating Engel curves for food. Am. Econ. Rev. 102, 1093–1117.
Apps, P., Rees, R., 1988. Taxation and the household. J. Pub. Econ. 35, 355–369.
Arneson, R., 1989. Equality and equal opportunity for welfare. Philos. Stud. 56, 77–93.
Arrow, K., 1951. Social Choice and Individual Values. Wiley, New York.
Arrow, K., 1973. Some ordinalist-utilitarian notes on Rawls's theory of justice. J. Philos. 70, 245–263.
Atkinson, A., 2005. Atkinson Review: Measurement of Government Output and Productivity for the National Accounts. Final Report. Palgrave Macmillan, Basingstoke.
Atkinson, A., Bourguignon, F., 1982. The comparison of multi-dimensioned distributions of economic status. Rev. Econ. Stud. 49, 183–201.
Atkinson, A., Bourguignon, F., 1987. Income distribution and differences in needs. In: Feiwel, G.F. (Ed.), Arrow and the Foundation of the Theory of Economic Policy. Macmillan, London, pp. 350–370.
Barbera, S., Bossert, W., Pattanaik, P., 1998. Ranking sets of objects. In: Barbera, S., Hammond, P., Seidl, C. (Eds.), Handbook of Utility Theory. Kluwer Academic Press, Dordrecht, Chapter 17.
Bargain, O., Decoster, A., Dolls, M., Neumann, D., Peichl, A., Siegloch, S., 2013. Welfare, labor supply and heterogeneous preferences: evidence for Europe and the US. Soc. Choice Welfare 41, 789–817.
Basu, K., 1987. Achievements, capabilities and the concept of well-being. Soc. Choice Welfare 4, 69–76.
Basu, K., Lopez-Calva, L., 2011. Functionings and capabilities. In: Arrow, K., Sen, A., Suzumura, K. (Eds.), Handbook of Social Choice and Welfare. Elsevier, Amsterdam, pp. 153–187.
Becker, G., Philipson, T., Soares, R., 2005. The quantity and quality of life and the evolution of world inequality. Am. Econ. Rev. 95, 277–291.
Benjamin, D., Heffetz, O., Kimball, M., Rees-Jones, A., 2012. What do you think would make you happier? What do you think you would choose? Am. Econ. Rev. 102, 2083–2110.
Bernheim, D., 2009. Behavioral welfare economics. J. Eur. Econ. Assoc. 7, 267–319.
Bernheim, D., Rangel, A., 2009. Beyond revealed preference: choice theoretic foundations for behavioral welfare economics. Q. J. Econ. 124, 51–104.
Blackorby, C., Donaldson, D., 1988. Money metric utility: a harmless normalization? J. Econ. Theory 46, 120–129.
Blackorby, C., Donaldson, D., 1993. Adult-equivalence scales and the economic implementation of interpersonal comparisons of well-being. Soc. Choice Welfare 10, 335–361.
Blanchflower, D., Oswald, A., 2004. Well-being over time in Britain and the USA. J. Pub. Econ. 88, 1359–1386.
Bourguignon, F., 1989. Family size and social utility: income distribution dominance criteria. J. Econ. 42, 67–80.

Bourguignon, F., Rogers, H., 2007. Distributional effects of educational improvements: are we using the wrong model? Econ. Educ. Rev. 26, 735–746.

Browning, M., Chiappori, P.-A., Lewbel, A., 2013. Estimating consumption economies of scale, adult equivalence scales, and household bargaining power. Rev. Econ. Stud. 80, 1267–1303.

Brun, B., Tungodden, B., 2004. Non-welfaristic theories of justice: is the "intersection approach" a solution to the indexing impasse? Soc. Choice Welfare 22, 49–60.

Carson, R., 2012. Contingent valuation: a practical alternative when prices aren't available. J. Econ. Perspect. 26, 27–42.

Cherchye, L., De Rock, B., Vermeulen, F., 2012. Economic well-being and poverty among the elderly: an analysis based on a collective consumption model. Eur. Econ. Rev. 56, 985–1000.

Cherchye, L., De Rock, B., Lewbel, A., Vermeulen, F., 2013. Sharing Rule Identification for General Collective Consumption Models, Mimeo.

Chiappori, P.-A., 1988. Rational household labor supply. Econometrica 56, 63–89.

Clark, D., 2005. Sen's capability approach and the many spaces of human well-being. J. Dev. Stud. 41, 1339–1368.

Clark, A., Oswald, A., 2002. A simple statistical method for measuring how life events affect happiness. Int. J. Epidemiol. 31, 1139–1144.

Cohen, G., 1989. On the currency of egalitarian justice. Ethics 99, 906–944.

Crawford, I., Neary, P., 2008. Testing for a reference consumer in international comparisons of living standards. Am. Econ. Rev. 98, 1731–1732.

Currie, J., Gahvari, F., 2008. Transfers in cash and in-kind: theory meets the data. J. Econ. Lit. 46, 333–383.

d'Aspremont, C., Gevers, L., 1977. Equity and the informational basis of collective choice. Rev. Econ. Stud. 44, 199–209.

d'Aspremont, C., Gevers, L., 2002. Social welfare functionals and interpersonal comparability. In: Arrow, K., Sen, A., Suzumura, K. (Eds.), In: Handbook of Social Choice and Welfare, vol. 1. Elsevier, Amsterdam, pp. 459–541.

Dardanoni, V., 1995. On multidimensional inequality measurement. In: Dagum, C., Lemmi, A. (Eds.), In: Research on Economic Inequality, vol. 6. JAI Press, London, pp. 201–207.

Deaton, A., 2008. Income, health, and well-being around the world: evidence from the Gallup World Poll. J. Econ. Perspect. 22, 53–72.

Deaton, A., 2010. Price indexes, inequality, and the measurement of world poverty. Am. Econ. Rev. 100, 5–34.

Deaton, A., Heston, A., 2010. Understanding PPPs and PPP-based national accounts. Am. Econ. J. Macroecon. 2, 1–35.

Deaton, A., Muellbauer, J., 1980. Economics and Consumer Behaviour. Cambridge University Press, Cambridge.

Decancq, K., 2013. Copula-based measurement of dependence between dimensions of well-being. Oxf. Econ. Pap. 66 (3), 681–701.

Decancq, K., Lugo, M.A., 2012. Inequality of well-being: a multidimensional approach. Economica 79, 721–746.

Decancq, K., Lugo, M.A., 2013. Weights in multidimensional indices of well-being: an overview. Econom. Rev. 32, 7–34.

Decancq, K., Ooghe, E., 2010. Has the world moved forward? A robust multidimensional evaluation. Econ. Lett. 107, 266–269.

Decancq, K., Schokkaert, E., 2013. Beyond GDP: Measuring Social Progress in Europe. KU Leuven, Leuven. Mimeo.

Decancq, K., Decoster, A., Schokkaert, E., 2009. The evolution of world inequality in well-being. World Dev. 37, 11–25.

Decoster, A., Haan, P., 2013. Empirical welfare analysis in random utility models of labour supply. Int. Tax Pub. Finance, forthcoming.

Della Vigna, S., 2009. Psychology and economics: evidence from the field. J. Econ. Lit. 47, 315–372.

de Ree, J., Alessie, R., Pradhan, M., 2013. The price and utility dependence of equivalence scales: evidence from Indonesia. J. Pub. Econ. 97, 272–281.

Diener, E., Suh, E., Lucas, R., Smith, H., 1999. Subjective well-being: three decades of progress. Psychol. Bull. 125, 276–302.
Diener, E., Helliwell, J., Kahneman, D. (Eds.), 2010. International Differences in Wellbeing. Oxford University Press, Oxford.
Diewert, W., 1976. Exact and superlative index numbers. J. Econ. 4, 115–145.
Diewert, W., 1992a. Exact and superlative welfare change indicators. Econ. Inq. 30, 565–583.
Diewert, W., 1992b. Fisher ideal output, input, and productivity indexes revisited. J. Prod. Anal. 3, 211–248.
Donaldson, D., 1992. On the aggregation of money measures of well-being in applied welfare economics. J. Agric. Resour. Econ. 17, 88–102.
Donaldson, D., Pendakur, K., 2003. Equivalent-expenditure functions and expenditure-dependent equivalence scales. J. Pub. Econ. 88, 175–208.
Dunbar, G., Lewbel, A., Pendakur, K., 2013. Children's resources in collective households: identification, estimation, and an application to child poverty in Malawi. Am. Econ. Rev. 103, 438–471.
Dutta, I., Foster, J., 2013. Inequality of happiness in the U.S.: 1972–2010. Rev. Income Wealth 59, 393–415.
Dutta, I., Pattanaik, P., Xu, Y., 2003. On measuring deprivation and the standard of living in a multidimensional framework on the basis of aggregate data. Economica 70, 197–221.
Dworkin, R., 1981a. What is equality? Part 1: equality of welfare. Philos. Pub. Aff. 10, 185–246.
Dworkin, R., 1981b. What is equality? Part 2: equality of resources. Philos. Pub. Aff. 10, 283–345.
Dworkin, R., 2000. Sovereign Virtue. Cambridge University Press, Cambridge.
Easterlin, R., 1974. Does economic growth improve the human lot? Some empirical evidence. In: David, P., Reder, M. (Eds.), Nations and Households in Economic Growth: Essays in Honor of Moses Abramovitz. Academic Press, New York, pp. 89–125.
European Statistical System, 2011. Sponsorshop Group on Measuring Progress, Wellbeing and Sustainable Development: Final Report. http://epp.eurostat.ec.europa.eu/portal/page/portal/pgp_ess/0_DOCS/estat/SpG_progress_wellbeing_report_after_ESSC_adoption_22Nov1.pdf.
Fisher, I., 1922. The Making of Index Numbers. Houghton Mifflin, Boston.
Fisher, G.M., 1997. The development of the Orshansky poverty thresholds and their subsequent history as the official U.S. Poverty Measure. Mimeo on, http://www.census.gov/hhes/povmeas/publications/orshansky.html.
Fleurbaey, M., 2003. On the informational basis of social choice. Soc. Choice Welfare 21, 347–384.
Fleurbaey, M., 2006a. Capabilities, functionings and refined functionings. J. Hum. Dev. 7, 299–310.
Fleurbaey, M., 2006b. Social welfare, priority to the worst-off and the dimensions of individual well-being. In: Farina, F., Savaglio, E. (Eds.), Inequality and Economic Integration. Routledge, London, pp. 225–268.
Fleurbaey, M., 2007. Social choice and the indexing dilemma. Soc. Choice Welfare 29, 633–648.
Fleurbaey, M., 2008. Fairness, Responsibility and Welfare. Oxford University Press, Oxford.
Fleurbaey, M., 2009. Equality of functionings. Социальные проблемы, http://socprob.ru/2009/equality-of-functionings-mark-flerbe.html.
Fleurbaey, M., Blanchet, D., 2013. Beyond GDP. Measuring Welfare and Assessing Sustainability. Oxford University Press, Oxford.
Fleurbaey, M., Gaulier, G., 2009. International comparisons of living standards by equivalent incomes. Scand. J. Econ. 111, 597–624.
Fleurbaey, M., Maniquet, F., 2011. A Theory of Fairness and Social Welfare. Cambridge University Press, Cambridge.
Fleurbaey, M., Mongin, P., 2005. The news of the death of welfare economics is greatly exaggerated. Soc. Choice Welfare 25, 381–418.
Fleurbaey, M., Schokkaert, E., 2012. Inequity in health and health care. In: Barros, P., McGuire, T., Pauly, M. (Eds.), Handbook of Health Economics, vol. 2. Elsevier, New York, pp. 1003–1092.
Fleurbaey, M., Schokkaert, E., 2013. Behavioral welfare economics and redistribution. Am. Econ. J. Microecon. 5, 180–205.
Fleurbaey, M., Tadenuma, K., 2007. Do irrelevant commodities matter? Econometrica 75, 1143–1174.
Fleurbaey, M., Trannoy, A., 2003. The impossibility of a Paretian egalitarian. Soc. Choice Welfare 21, 243–263.

Fleurbaey, M., Hagneré, C., Trannoy, A., 2003. Welfare comparisons with bounded equivalence scales. J. Econ. Theory 110, 309–336.
Fleurbaey, M., Schokkaert, E., Decancq, K., 2009. What Good Is Happiness?, CORE: Discussion Paper 2009/17.
Fleurbaey, M., Luchini, S., Schokkaert, E., Van de Voorde, C., 2012. Evaluation des politiques de santé: pour une prise en compte équitable des interest des populations. Econ. Stat. 455–456, 11–36.
Fleurbaey, M., Luchini, S., Muller, C., Schokkaert, E., 2013. Equivalent income and the economic evaluation of health care. Health Econ. 22, 711–729.
Foster, J., 2010. Freedom, opportunity, and well-being. In: Arrow, K.J., Sen, A.K., Suzumura, K. (Eds.), Handbook of Social Choice and Welfare, vol. 2. North Holland, Amsterdam, pp. 687–728.
Foster, J., McGillivray, M., Seth, S., 2013. Composite indices. Rank robustness, statistical association and redundancy. Econom. Surv. 32, 35–56.
Frey, B., Stutzer, A., 2002. What can economists learn from happiness research? J. Econ. Lit. 40, 402–435.
Gadrey, J., Jany-Catrice, F., 2006. The New Indicators of Well-Being and Development. Macmillan, London.
Gajdos, T., Weymark, J., 2005. Multidimensional generalized Gini indices. Econ. Theory 26, 471–496.
Graham, C., 2009. Happiness Around the World: The Paradox of Happy Peasants and Miserable Millionaires. Oxford University Press, Oxford.
Gruber, J., Mullainathan, S., 2005. Do cigarette taxes make smokers happier? Adv. Econ. Anal. Pol. 5, Article 4. Www.Bepress.Com.
Hausman, J., 2012. Contingent valuation: from dubious to hopeless. J. Econ. Perspect. 26, 43–56.
Hausman, D., McPherson, M., 2009. Preference satisfaction and welfare economics. Econ. Philos. 25, 1–25.
Helliwell, J., Barrington-Leigh, C., Harris, A., Huang, H., 2010. International evidence on the social context of well-being. In: Diener, E., Helliwell, J., Kahneman, D. (Eds.), International Differences in Well-Being. Oxford University Press, Oxford.
Jones, C., Klenow, P., 2010. Beyond GDP? Welfare Across Countries and Time. NBER: Working Paper 16352.
Jones, A., O'Donnell, O., 1995. Equivalence scales and the costs of disability. J. Pub. Econ. 56, 273–289.
Kahneman, D., Krueger, A., 2006. Developments in the measurement of subjective well-being. J. Econ. Perspect. 20, 3–24.
Kahneman, D., Sugden, R., 2005. Experienced utility as a standard of policy evaluation. Environ. Resour. Econ. 32, 161–181.
Kahneman, D., Wakker, P., Sarin, R., 1997. Back to Bentham? Explorations of experienced utility. Q. J. Econ. 112, 375–406.
Kahneman, D., Krueger, A., Schkade, D., Schwarz, N., Stone, A., 2004. Toward national well-being accounts. Am. Econ. Rev. 94, 429–434, Papers and Proceedings.
Kapteyn, A., 1994. The measurement of household cost functions: revealed preference versus subjective measures. J. Popul. Econ. 4, 333–350.
Kapteyn, A., Van Praag, B., 1976. A new approach to the construction of family equivalence scales. Eur. Econ. Rev. 7, 313–335.
Kimball, M., Willis, R., 2006. Utility and Happiness. Mimeo.
King, M., 1983. Welfare analysis of tax reforms using household data. J. Pub. Econ. 23, 183–214.
Kolm, S.-Chr., 1977. Multidimensional egalitarianisms. Q. J. Econ. 91, 1–13.
Koshevoy, G., 1995. Multivariate Lorenz majorization. Soc. Choice Welfare 12, 93–102.
Koulovatianos, C., Schröder, C., Schmidt, U., 2005. On the income dependence of equivalence scales. J. Pub. Econ. 89, 967–996.
Layard, R., 2005. Happiness: Lessons from a New Science. Allan Lane, London.
Lelli, S., 2005. Using functionings to estimate equivalence scales. Rev. Income Wealth 51, 255–284.
Levinson, A., 2012. Valuing public goods using happiness data: the case of air quality. J. Pub. Econ. 96, 869–880.
Lewbel, A., 1989. Household equivalence scales and welfare comparisons. J. Pub. Econ. 39, 377–391.
Lewbel, A., 2003. Calculating compensation in cases of wrongful death. J. Econ. 113, 115–128.
Loewenstein, G., Ubel, P., 2008. Hedonic adaptation and the role of decision and experienced utility in public policy. J. Pub. Econ. 92, 1795–1810.

Manser, M., Brown, M., 1980. Marriage and household-decision making: a bargaining analysis. Int. Econ. Rev. 21, 31–44.
Marical, F., Mira d'Ercole, M., Vaalavuo, M., Verbist, G., 2008. Publicly Provided Services and the Distribution of Households' Economic Resources. OECD Economic Studies No. 44, 2008/1.
Marshall, A., Olkin, I., 1979. Inequalities: Theory of Majorization and Its Applications. Academic Press, New York.
McElroy, M., Horney, J., 1981. Nash-bargained household decisions: towards a generalization of the theory of demand. Int. Econ. Rev. 22, 333–349.
Musgrave, R.A., 1959. The Theory of Public Finance. McGraw-Hill, New York.
Neary, P., 2004. Rationalizing the Penn world table: true multilateral indices for international comparisons of real income. Am. Econ. Rev. 94, 1411–1428.
Nussbaum, M., 2000. Women and Human Development: The Capabilities Approach. Cambridge University Press, Cambridge.
Nussbaum, M., 2006. Frontiers of Justice: Disability, Nationality, Species Membership. Harvard University Press, Cambridge, MA.
Nussbaum, M., 2008. Who is the happy warrior? Philosophy poses questions to psychology. J. Legal Stud. 37, S81–S113.
Nussbaum, M., 2011. Creating Capabilities: The Human Development Approach. Belknap Press of Harvard University Press, Cambridge, MA.
Nussbaum, M., Sen, A. (Eds.), 1993. The Quality of Life. Clarendon Press, Oxford.
OECD, 2008. Growing Unequal? Income Distribution and Poverty in OECD Countries. OECD Publishing, Paris.
OECD, 2011. How's Life? Measuring Well-Being. OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264121164-en.
Ooghe, E., Lambert, P., 2006. On bounded dominance criteria. Math. Soc. Sci. 52, 15–30.
Oswald, A., 1997. Happiness and economic performance. Econ. J. 107, 1815–1831.
Oswald, A., Powdthavee, N., 2008. Does happiness adapt? A longitudinal study of disability with implications for economists and judges. J. Pub. Econ. 92, 1061–1077.
Pattanaik, P., Xu, Y., 1990. On ranking opportunity sets in terms of freedom of choice. Rech. Econ. Louvain 56, 383–390.
Pattanaik, P., Xu, Y., 2007. Minimal relativism, dominance, and standard of living comparisons based on functionings. Oxf. Econ. Pap. 59, 354–374.
Pattanaik, P., Xu, Y., 2009. Conceptions of individual rights and freedom in welfare economics: a re-examination. In: Gotoh, R., Dumouchel, P. (Eds.), Against Injustice: The New Economics of Amartya Sen. Cambridge University Press, Cambridge, pp. 187–218.
Pattanaik, P., Xu, Y., 2012. The Ethical Bases of Public Policies: A Conceptual Framework, Mimeo.
Pattanaik, P., Reddy, S., Xu, Y., 2012. On measuring deprivation and living standards of societies in a multi-attribute framework. Oxf. Econ. Pap. 64, 43–56.
Paulus, A., Sutherland, H., Tsakloglou, P., 2010. The distributional impact of in-kind public benefits in European countries. J. Policy Anal. Manage. 29, 243–266.
Pogge, T., 2002. Can the capability approach be justified? Philos. Top. 30, 167–228.
Pollak, R., Wales, T., 1979. Welfare comparisons and equivalence scales. Am. Econ. Rev. 62, 216–221, Papers and Proceedings.
Pont, M., 2008. Improvements to the measurement of government output in the national accounts. Econ. Labour Mark. Rev. 2, 17–22.
Quiggin, J., van Veelen, M., 2007. Multilateral indices: conflicting approaches? Rev. Income Wealth 53, 372–378.
Radner, D.B., 1997. Noncash income, equivalence scales, and the measurement of economic well-being. Rev. Income Wealth 43, 71–88.
Ravallion, M., 2012. Troubling tradeoffs in the human development index. J. Dev. Econ. 99, 201–209.
Rawls, J., 1971. A Theory of Justice. Harvard University Press, Cambridge, MA.
Rawls, J., 1982. Social unity and primary goods. In: Sen, A., Williams, B. (Eds.), Utilitarianism and Beyond. Cambridge University Press, Cambridge, pp. 159–185.

Rayo, L., Becker, G., 2007. Habits, peers and happiness: an evolutionary perspective. Am. Econ. Rev. 97, 487–491. Papers and Proceedings.
Robbins, L., 1938. Interpersonal comparisons of utility: a comment. Econ. J. 48, 635–641.
Robeyns, I., 2003. Sen's capability approach and gender inequality: selecting relevant capabilities. Fem. Econ. 9, 61–92.
Robeyns, I., 2006. The capability theory in practice. J. Polit. Philos. 14, 351–376.
Roemer, J., 1996. Theories of Distributive Justice. Harvard University Press, Cambridge.
Roemer, J., 1998. Equality of Opportunity. Harvard University Press, Cambridge, MA.
Samuelson, P., 1974. Complementarity: an essay on the 40th anniversary of the Hicks–Allen revolution in demand theory. J. Econ. Lit. 12, 1255–1289.
Samuelson, P., Swamy, S., 1974. Invariant economic index numbers and canonical duality: survey and synthesis. Am. Econ. Rev. 64, 566–593.
Scanlon, T., 1975. Preference and urgency. J. Philos. 72, 655–669.
Schokkaert, E., Van Ootegem, L., Verhofstadt, E., 2011. Preferences and subjective job satisfaction: measuring well-being on the job for policy evaluation. CESifo Econ. Stud. 57, 683–714.
Sen, A., 1970. Collective Choice and Social Welfare. Holden Day, San Francisco.
Sen, A., 1980. Equality of what? In: Sen, A. (Ed.), Choice, Welfare and Measurement. Blackwell, Oxford, pp. 353–369.
Sen, A., 1984. The living standard. Oxf. Econ. Pap. 36, 74–90.
Sen, A., 1985. Commodities and Capabilities. North-Holland, Amsterdam.
Sen, A., 1992. Inequality Re-Examined. Clarendon Press, Oxford.
Sen, A., 1999. Development as Freedom. Knopf, New York.
Sen, A., 2004. Capabilities, lists, and public reason: continuing the conversation. Fem. Econ. 10, 77–80.
Sen, A., 2009. The Idea of Justice. Allen Lane, London.
Sen, A., Williams, B. (Eds.), 1982. Utilitarianism and Beyond. Cambridge University Press, Cambridge.
Sen, A., Muellbauer, J., Kanbur, R., Hart, K., Williams, B., 1987. The Standard of Living. Cambridge University Press, Cambridge.
Slesnick, D.T., 1996. Consumption and poverty: how effective are in-kind transfers? Econ. J. 106, 1527–1545.
Smeeding, T., 1977. The antipoverty effectiveness of in-kind transfers. J. Hum. Resour. 12, 360–378.
Smeeding, T., 1982. Alternative Methods for Valuing In-Kind Transfer Benefits and Measuring Their Impact on Poverty: Technical Report 50. U.S. Bureau of the Census, Washington, DC.
Smeeding, T., Coder Saunders, P., Jenkins, S., Fritzell, J., Haagenaars, A.J.M., Hauser, R., Wolfson, M., 1993. Poverty, inequality, and family living standards impact across seven nations: the effect of noncash subsidies for health, education and housing. Rev. Income Wealth 39, 229–256.
Smolensky, E., Stiefel, L., Schmundt, M., 1977. Adding in-kind transfers to the personal income and outlay account: implications for the size distribution of income. In: Juster, T.F. (Ed.), Distribution of Economic Well-Being. NBER, Cambridge, MA, pp. 9–50.
Sprumont, Y., 2012. Resource egalitarianism with a dash of efficiency. J. Econ. Theory 147, 1602–1613.
Stevenson, B., Wolfers, J., 2008. Economic growth and subjective well-being: reassessing the Easterlin paradox. Brook. Pap. Econ. Act. 1, 1–97.
Stiglitz, J., Sen, A., Fitoussi, J.-P., 2009. Report by the Commission on the Measurement of Economic Performance and Social Progress, Paris.
Stutzer, A., Frey, B., 2008. Stress that doesn't pay: the commuting paradox. Scand. J. Econ. 110, 339–366.
Sugden, R., 1993. Welfare, resources and capabilities: a review of "Inequality Reexamined" by Amartya Sen. J. Econ. Lit. 31, 1947–1962.
Trannoy, A., 2006. Multidimensional egalitarianism and the dominance approach: a lost paradise? In: Farina, F., Savaglio, E. (Eds.), Inequality and Economic Integration. Routledge, London, pp. 284–302.
Tsui, K.-Y., 1995. Multidimensional generalizations of the relative and absolute inequality indices: the Atkinson–Kolm–Sen-approach. J. Econ. Theory 67, 251–265.
UNDP, 1990. Human Development Report 1990. Oxford University Press, New York.
UNDP, 2010. The Real Wealth of Nations: Pathways to Human Development. Oxford University Press, New York.

Van Praag, B., 1971. The welfare function of income in Belgium: an empirical investigation. Eur. Econ. Rev. 2, 337–369.

Van Praag, B., Ferrer-i-Carbonell, A., 2007. Happiness Quantified: A Satisfaction Calculus Approach. Oxford University Press, Oxford.

Van Praag, B., Ferrer-i-Carbonell, A., 2009. Inequality and happiness. In: Salverda, W., Nolan, B., Smeeding, T. (Eds.), Oxford Handbook of Economic Inequality. Oxford University Press, Oxford, Chapter 17.

Van Praag, B., van der Sar, N., 1988. Household cost functions and equivalence scales. J. Hum. Resour. 23, 193–210.

van Veelen, M., 2002. An impossibility theorem concerning multilateral international comparisons of volume. Econometrica 70, 369–375.

van Veelen, M., van der Weide, R., 2008. A note on different approaches to index number theory. Am. Econ. Rev. 98, 1722–1730.

Verbist, G., Förster, M.F., Vaalavuo, M., 2012. The Impact of Publicly Provided Services on the Distribution of Resources: Review of New Results and Methods, OECD Social, Employment and Migration Working Papers, No. 130.

Weymark, J., 2006. The normative approach to the measurement of multidimensional inequality. In: Farina, F., Savaglio, E. (Eds.), Inequality and Economic Integration. Routledge, London, pp. 303–328.

Weymark, J., 2013. Conundrums for Nonconsequentialists, Vanderbilt University Department of Economics Working Papers 13-00010.

CHAPTER 3

Multidimensional Poverty and Inequality

Rolf Aaberge*, Andrea Brandolini†
*Statistics Norway, Research Department and University of Oslo, Department of Economics, ESOP, Oslo, Norway
†Bank of Italy, Directorate General for Economics, Statistics, and Research, Rome, Italy

Contents

Abstract

This chapter examines different approaches to the measurement of multidimensional inequality and poverty. It first outlines three aspects preliminary to any multidimensional study: the selection of the relevant dimensions, the indicators used to measure them, and the procedures for their weighting. It

Handbook of Income Distribution, Volume 2A
ISSN 1574-0056, http://dx.doi.org/10.1016/B978-0-444-59428-0.00004-7

then considers the counting approach and the axiomatic treatment in poverty measurement. Finally, it reviews the axiomatic approach to inequality analysis. The chapter also provides a selective review of the rapidly growing theoretical literature with the twofold aim of highlighting areas for future research and offering some guidance on how to use multidimensional methods in empirical and policy-oriented applications.

Keywords

Inequality, Poverty, Deprivation, Multidimensional well-being, Capabilities and functionings

JEL Classification Codes

D3, D63, I30, I32

3.1. INTRODUCTION

Few people would question that well-being results from many different attributes of human life, and the level of income, or expenditure, is only a crude proxy of the quality of living that a person enjoys.[1] Should we then account for the multiple facets of well-being in the social evaluation of inequality and poverty? If so, how can we do it?

Acknowledging the multidimensional nature of well-being does not necessarily imply that the social evaluation must also be multidimensional. Some argue that a single variable can still subsume all various dimensions of well-being. This is typically the case in the utilitarian approach, which employs a single indicator represented by "utility," or the level of well-being as assessed by individuals. Individuals themselves reduce the vector $\boldsymbol{x}$ of the different constituents of well-being to the level of utility $u(\boldsymbol{x})$. The social evaluation may then consider estimated utility levels as revealed by individuals, either directly through their answers to questions on subjective well-being and life satisfaction, as in the happiness literature,[2] or indirectly through their consumption patterns, as suggested by Jorgenson and Slesnick (1984a,b). Apart from requiring analytical restrictions (e.g., shape of indirect utility functions, integrability of demand functions), these approaches run into the difficulty that individual utilities must be assumed to be interpersonally comparable. Alternatively, the reduction of multiple dimensions to a single indicator can be

[1] Throughout the chapter, we interchangeably use terms such as "well-being", "quality of life," and "standard of living," without adopting any precise definition, except for the recognition of their multidimensional nature. The ensuing ambiguity is not a problem for our presentation, but it might be in a different context. For a discussion of this point, see the exchange between Williams (1987), Sen (1987), and Sen (1993). Likewise, we use terms such as "attributes," "dimensions," and "domains" to indicate the components of a multivariate notion of deprivation or well-being, although we acknowledge that, in certain areas of the literature on social indicators, they may be used to indicate different concepts.

[2] Well before the recent surge of interest in happiness research among economists, the "Leyden approach" to the measurement of poverty proposed exploiting the information on people's subjective evaluation of their own economic condition to identify poverty thresholds. See, for instance, Goedhart et al. (1977), van Praag et al. (1980), Danziger et al. (1984), van Praag et al. (2003), and van Praag and Ferrer-i-Carbonell (2008).

considered to be carried out by a social evaluator. This composite indicator would then represent a "utility-like function of all the attributes received," as put by Maasoumi (1986, p. 991), to which standard univariate techniques could be applied. Maasoumi suggests applying information theory to find the utility-like function with a distribution as close as possible to the distributions of the constituent attributes, but other approaches can lead to the definition of analogous individual-level functions. The common practice of adjusting household income for the household size and the age of its members by an equivalence scale is another example of this type of multidimensional analysis, for which command over resources (income) and individual needs (varying by age and living arrangements) are the two dimensions reputed to be relevant in assessing well-being. The chosen equivalence scale is assumed to represent the preferences of the social evaluator.

At the opposite extreme are those who argue, on philosophical or practical grounds, that dimensions must be kept distinct in the social evaluation. If well-being domains are characterized by specific criteria and arrangements, some might adhere to Walzer's (1983, p. 19) view of "complex equality" whereby "no citizen's standing in one sphere or with regard to one social good can be undercut by his standing in some other sphere, with regard to some other good." If inequalities in certain domains (e.g., basic life necessities or health) are less acceptable than they are in others (e.g., luxury goods), it might be justifiable to adopt a piecemeal approach informed by the "specific egalitarianism" advocated by Tobin (1970).[3] The intrinsic incommensurability of domains may then imply that "no simple ordered indicator of level of living can be constructed, either on an individual or on an aggregate level," as asserted by Erikson (1993, p. 75) in summarizing the Swedish approach to welfare research. Or, the need to avoid the "*ad hoc* aggregation" and the unexplained trade-offs between domains, which are implicit in any composite or "mashup" index, might advise us "to derive the best measure possible for each of a logically defensible set of grouped dimensions—such as 'income poverty,' 'health poverty' and 'education poverty'" (Ravallion, 2011a, p. 240; see also Ravallion, 2012a). In all these cases, the recognition of the inherent autonomy of each dimension, however motivated, leads to a piecewise social judgement that does not need any unitary measurement of human well-being. The elements of the vector $\boldsymbol{x}$ of the attributes of well-being are examined one by one, without attempting to reduce complexity by a summary index. It is the "dashboard" approach. The straightforwardness of this strategy is appealing, but it is tempered by the difficulty of drawing a synthetic picture, especially in the presence of a rich information set.

[3] Slesnick (1989) assesses how pursuing equalization in separate domains affects the inequality of overall utility (specific egalitarianism vs. general egalitarianism) by comparing the inequality of main consumption components with the inequality of total expenditure.

There are reasons to take an intermediate position between these two extremes, however. This may be because the above-described conditions for reducing well-being to a single variable may not hold: we might differ in our views about the appropriate equivalence scale or the weights to be placed on different goods, we might not have access to individual well-being measures, or we might reject the individual valuations altogether. Or, we may worry that the inequalities in different spheres accumulate and that the combination of multiple deprivations makes life much harder than just the sum of such deprivations. In these cases, we may need a social evaluation of poverty and inequality that is multidimensional and accounts for the joint distributions of all the elements of the vector $\boldsymbol{x}$ of well-being attributes.

Our aim in this chapter is to explore this intermediate route. We do not argue further whether we should or should not have a multidimensional social evaluation. We take it for granted, and we concentrate on how we can carry it out in a sound way. More precisely, we examine the analytical and ethical foundations of methods for the multidimensional measurement of inequality and poverty, whether it be for descriptive, normative, or policy-making purposes. All these methods require numerous arbitrary, and hence debatable, assumptions. Elucidating their foundations helps us to unveil these assumptions and understand their normative content. Taking this perspective, we pay little attention to the many multivariate techniques that have been developed in statistics and efficiency analysis. They provide valuable information, but their aggregation of multiple attributes is based on empirically observed patterns of association among the variables, and, thus, it lacks any clear ethical interpretation. We may legitimately hesitate to entrust a mathematical algorithm with an essentially normative task such as deriving an index of well-being.

The theoretical literature on the multidimensional measurement of inequality and poverty has been growing very rapidly in the last quarter of a century, and it is still far from consolidation. Rather than engaging in a systematic rationalization of this literature, we provide a selective reading of it with the twofold objective of, first, identifying areas worthy of further investigation and, second, offering some guidance on how to use the rich and sophisticated machinery now available for empirical and policy-oriented applications. As the multidimensional view of well-being has gained momentum in the policy discourse, its practical implementation has turned into an active battlefield where contenders passionately argue for opposing approaches—a good example being the Forum on multidimensional poverty in the 2011 volume of the *Journal of Economic Inequality* (see Lustig, 2011, for an introduction). We attempt to provide a balanced account of alternative positions, as well as their strengths and weaknesses.

The chapter is divided into three parts, plus a closing section. In the next section, we briefly review three questions that are preliminary to any multidimensional analysis of well-being: the selection of the relevant dimensions, the indicators used to measure them, and the procedures for their weighting. These questions are theoretically intriguing and of considerable importance in empirical analyses, but we only outline their main features. Importantly, the choice made with regard to these issues may condition the analytical methods reviewed later. For instance, the fact that many variables used in multidimensional poverty analysis are dichotomous suggests paying particular attention to methods based on counting deprivations. The assumption that inequality does not change after proportionate variations of the variable under examination (scale invariance) may be reasonable for income, but much less so for life expectancy, impinging on the axiomatic measurement of multidimensional inequality. We then move to the core of the chapter: the methods for the multivariate analysis of poverty, in Section 3.3, and of inequality, in Section 3.4. In the remainder of this introduction, we offer a brief account of the historical developments in the research summarized in these two sections, while providing a tour of the main themes discussed in the chapter.

3.1.1 Historical Developments and Main Themes

The multidimensional literature in economics began with seminal articles by Kolm (1977) and by Atkinson and Bourguignon (1982) on the dominance conditions for ranking multivariate distributions. A few years later, Atkinson and Bourguignon (1987) developed sequential dominance criteria for the bivariate space of income and household composition. Their aim was to impose weaker assumptions on social preferences than those implicit in the standard method of constructing equivalent incomes. Whereas the standard approach entails specifying how much a family type is needier than another one, sequential dominance criteria only require ranking family types in terms of needs, although at the cost of obtaining an incomplete ordering. This application paved the way for a specific and fertile strand of research which focuses on the possibility that one attribute (e.g., income) can be used to compensate for another nontransferable attribute (e.g., needs, health).

With the partial exception of Maasoumi (1986, 1989), who recasted multidimensional analysis into a unidimensional space by means of a utility-like function, it was not until the mid-1990s that Tsui (1995, 1999) moved on to the axiomatic approach to inequality indices in order to achieve complete orderings. The bases of the axiomatic analysis of partial and complete poverty orderings were laid down at about the same time by Chakravarty et al. (1998), Bourguignon and Chakravarty (1999, 2003, 2009), and Tsui (2002). Notably, multidimensional indices of inequality and poverty associate real numbers to each multivariate distribution, as does the univariate analysis of a composite well-being indicator, but with the important difference that they do not need to go through the aggregation of well-being attributes at the individual level. Thus, multidimensional poverty indices allow for separate thresholds for each attribute, while a utility-like indicator usually has a single threshold in the space of well-being. The trade-offs between the attributes that are built into the utility-like indicator used in the latter approach follow from the weighting structure of dimensions in the former approach.

At the turn of the twentieth century, the literature on multidimensional poverty and inequality was still in its infancy. The first volume of this *Handbook* (Atkinson and Bourguignon, 2000) did not feature any specific chapter on the topic, and the comprehensive analytical chapter on the measurement of inequality by Cowell (2000) devoted only three pages to multidimensional approaches. Ever since, the theoretical literature has grown conspicuously, however, and we can identify two main lines of research.

The first line devotes considerable effort to developing the axiomatic approach to both poverty and inequality measurement. Researchers delve into the different ways to model the patterns of association (correlation) between the variables, which is the single feature that distinguishes multidimensional from unidimensional analysis, and elaborate alternative axioms. They have also come to realize that a mechanical transposition of the properties typically adopted in the univariate analysis of income distribution may not be straightforward, and sometimes it may not even be appropriate. A case in point is the extension to life expectancy of the scale invariance property of inequality measures just mentioned. An even more cogent example is the Pigou–Dalton principle of transfers, a central tenet of income inequality measurement (Atkinson and Brandolini, forthcoming). This principle states that a mean-preserving transfer of income from a richer person to an (otherwise identical) poorer person decreases inequality. On the one hand, an interpersonal transfer might be unfeasible and even ethically debatable for a dimension such as the health status, despite being acceptable for income. On the other hand, the generalization of the principle to a multivariate framework is far from univocal, as explained in detail in Section 3.4.1.

The second line of research focuses on what Atkinson (2003, p. 51) labels the "counting approach." This multidimensional approach is at the same time the newest (in terms of theoretical elaboration) and the oldest (in terms of empirical practice). For example, the main poverty statistic adopted by a parliamentary commission of inquiry over destitution in Italy in the early 1950s was a weighted count of the number of households failing to achieve minimum levels of food consumption, clothing availability, and housing conditions (Cao-Pinna, 1953). Modern applied research on material deprivation owes much to the pioneering work by Townsend (1979) and Mack and Lansley (1985) in Britain.[4] Ever since the publication of their studies, it has had a huge

[4] Interestingly, Townsend's interest for elaborating a deprivation score was largely instrumental, being conceived as a way to reduce the arbitrariness of fixing income thresholds: "We assume that the deprivation index will not be correlated uniformly with total resources at the lower levels and that there will be a 'threshold' of resources below which deprivation will be marked" (Townsend, 1970, p. 29). There is by now an extensive literature on this subject. Some examples of studies for rich countries are Mayer and Jencks (1989), Federman et al. (1996), Nolan and Whelan (1996a,b, 2007, 2010, 2011), Whelan et al. (2001), Halleröd et al. (2006), Guio (2005), Cappellari and Jenkins (2007), Fusco and Dickes (2008), Fusco et al. (2010), and Figari (2012).

impact on the social policy debate in Ireland and the United Kingdom, and later in the European Union.[5] Nevertheless, we lack a full-fledged theoretical treatment of the normative basis of the counting approach. The recent work by Alkire and Foster (2011a,b) fills this gap, in part, by providing the axiomatic characterization of a family of multidimensional counting poverty indices. Yet, the difficulties illustrated by Atkinson (2003) in reconciling the counting approach with a social welfare approach are still unsettled. In our view, part of the problem may derive from defining welfare criteria in terms of the distributions of the underlying continuous variables rather than in terms of the distribution of deprivation scores, which is the key variable considered in the counting approach. The distribution of deprivation scores contains all the relevant information in the counting approach, which by construction implies neglecting levels of achievement in the original variables. Disagreement on this point, and on the implicit loss of information, might have some part in the recent controversies surrounding the counting approach.

The less developed analytical structure, in the face of the method's popularity in applied research, is the main reason for devoting a relatively larger space to the counting approach in this chapter. However, counting deprivations is also the simplest way to embed the association between individual-level dimensions into an overall index of deprivation. It is useful to illustrate two aspects of multidimensional measurement that recur throughout the chapter. The first is the order of aggregation. In the counting approach, the synthesis of available information begins with aggregation across single dimensions for each individual, and then across individuals. Inverting the order of aggregation by first computing the proportions of people suffering from deprivation in each dimension and then aggregating these proportions into a composite index of deprivation yields the same result only if the dimensions of well-being are "independent." If this is not the case, the composite index of deprivation misses the impact of cumulating failures in more than one dimension. The second aspect is the contrast between the "union criterion" and the "intersection criterion," which plays a

[5] Since 1997, the official poverty statistic adopted by the Irish government is "consistent poverty," which is the proportion of people who are both income-poor and deprived of two or more items considered essential for a basic standard of living (Social Inclusion Division, 2014). The British Child Poverty Act 2010 sets four policy targets, among which is a combined low income and material deprivation target (The Child Poverty Unit, 2014). One of the five European Union headline targets set by the Europe 2020 strategy for a smart, sustainable, and inclusive growth concerns the share of people "at risk of poverty or social exclusion" (European Commission 2010). This indicator combines income poverty, household joblessness, and severe material deprivation, with severe material deprivation occurring whenever a person lives in a household that cannot afford at least four out of nine amenities. For discussions of the use of indicators of material deprivation and, more generally, the multidimensional perspective adopted in the European Union's policy evaluation of social progress, see Atkinson et al. (2002), Marlier et al. (2007), Maquet and Stanton (2012), and Marlier et al. (2012).

fundamental role in the measurement of multidimensional poverty, as stressed by Atkinson (2003). The occurrence of deprivation in some dimensions does not necessarily entail a condition of overall poverty: we may define people to be poor when they are deprived in at least one dimension (union criterion) or in all dimensions (intersection criterion), or else in some fraction of the dimensions considered in the analysis. The choice of a critical number of dimensions to identify poverty status introduces an additional threshold relative to those already set for defining deprivation in each dimension, which is a central feature of the "dual cut-off" approach proposed by Alkire and Foster (2011a,b).

In Sections 3.3 and 3.4, we discuss first, the counting approach, then the axiomatic treatment of poverty, and finally, the axiomatic treatment of inequality. This sequence reflects the growing complexity of data requirements, rather than a chronological order. In this chapter we pay no attention to the assessment of data quality and the elaboration of inference tools, although they are admittedly two crucial issues in empirical analyses.

3.2. PRELIMINARIES: DIMENSIONS, INDICATORS, AND WEIGHTS

Three questions are preliminary to any discussion of the methods for the multivariate analysis of poverty and inequality: the selection of the relevant dimensions of well-being; the indicators used to measure people's achievements in these dimensions and the choice of deprivation thresholds in poverty analysis; and the weights assigned to each dimension. An in-depth examination of these issues is beyond the scope of this chapter, and our primary aim in this section is to highlight how these questions can influence the multivariate methods of analysis reviewed below. However, the actual solutions given to these questions may affect empirical findings and their substantive interpretation, and robustness and sensitivity exercises are advisable.

3.2.1 Selection of Dimensions

An established tradition of research in the study of deprivation postulates that we can better understand hardship by focusing on the individual's inability to consume socially perceived necessities because of a lack of economic resources, rather than focusing on income. Typically, this approach considers a battery of indicators concerning the ownership of durable goods; the possibility of carrying out certain activities, such as going out for a meal with friends; or the ability to cope with the payment of rent, mortgages, or utility bills. Material deprivation indicators have recently gained an official status in the monitoring of the social situation in the European Union, as well as in Ireland and the United Kingdom. The aim of the social evaluation may be broader than assessing material living conditions, however, and it may be concerned with "social exclusion."[6] According to Burchardt et al. (1999), social exclusion is associated with failures to achieve a reasonable living standard, a degree of security, an activity valued by others, some decision-making power, and the possibility of drawing support from relatives and friends. The variety of dimensions used to define the overall quality of life may be even larger. The Scandinavian approach to welfare, a long-established research program in Nordic countries, considers nine domains of human life: health and access to health care, employment and working conditions, economic resources, education and skills, family and social integration, housing, security of life and property, recreation and culture, and political resources (e.g., Erikson, 1993; Erikson and Uusitalo, 1986–87). Within the "capability approach," Nussbaum (2003) proposes a specific list of ten "central human capabilities": life; bodily health; bodily integrity; senses, imagination, and thought; emotions; practical reason; affiliation; other species; play; and control over one's environment. The Commission on the Measurement of Economic Performance and Social Progress, created at the beginning of 2008 under a French government initiative, identifies eight key dimensions: material living standards, health, education, personal activities including work, political voice and governance, social connections and relationships, environment (present and future conditions), and economic and physical insecurity (Stiglitz et al., 2009).

These examples illustrate the wide range and diversity of the domains considered in the multidimensional analysis of inequality and poverty. The choice of the dimensions that they include is mainly due to the advise of experts, possibly based on existing data, conventions, and statistical techniques.[7] It could also result from empirical evidence regarding citizen values, or it could be the product of a consultative process involving focus groups and representatives of the civil society or the public at large (Alkire, 2007). In all cases, the selection of poverty and income inequality indicators is a fundamental exercise, which has to blend theoretical rigor, political salience, empirical measurability, and data availability.

In this chapter, we simply assume that a predefined list of r attributes fully describes the well-being concept used in the analysis of poverty and inequality. We ignore all questions concerning the selection of attributes and refer the reader to Chapter 2 for a

[6] For a consideration of the somewhat elusive concept of social exclusion and its relationship with poverty, see Atkinson (1998). Ruggeri Laderchi et al. (2003) compare the empirical findings of social exclusion and capability approaches. Poggi (2007a,b) and Devicienti and Poggi (2011) empirically study the persistence of social exclusion, and Poggi and Ramos (2011) investigate the interdependency of the dimensions of social exclusion using stochastic epidemic models.

[7] For instance, Fusco and Dickes (2008) assume that poverty is a latent condition that can be identified by selecting the relevant domains from a set of deprivation indicators by applying a psychometric model.

comprehensive discussion.[8] Notice, however, that the nature of selected attributes may condition the definition of measurement tools. As noted in the introduction, we cannot mechanically export the Pigou–Dalton principle of transfers, which is central in income inequality analysis to other well-being dimensions, such as health (Bleichrodt and van Doorslaer, 2006), happiness (Kalmijn and Veenhoven, 2005), and literacy (Denny, 2002). Leaving aside the practical problem of how to transfer one unit of health from one person to another, we might doubt that imposing the principle of transfers in the health domain is ethically justified. We return to this issue in Section 3.4.1.

3.2.2 Indicators

The indicators used to measure people's achievements in the various dimensions are numerous and understandably have different measurement units. Incomes, wealth, and quantities consumed or purchased are continuous variables, but the number of durable goods owned and the frequency in the use of consumer services are discrete variables. Education can be measured by a categorical variable such as the highest school attainment of a person. Transforming it into the minimum number of years necessary to achieve each school level then provides an objective way to grade the various levels, but we might wonder whether a person who completed 14 years of school is really twice as well-educated as a person who only completed 7 years; moreover, only in a loose sense, can this type of transformed variable be interpreted as truly continuous. People's competencies and problem-solving capacity are increasingly assessed by complex exercises that produce literacy, numeracy, or skill scores generally normalized on a scale from 0 to 500. These scores are bounded, continuous, ordinal variables.[9] Individual health and physical status are measured with a host of indicators. Self-reported measures of health conditions are ordinal variables, but the information on the incidence of specific chronic illnesses is dichotomous; anthropometric indicators such as height, weight, and the body mass index are continuous variables. Subjective measures of well-being are typically collected by asking interviewees their personal degree of satisfaction on prefixed numerical scales or verbal rating scales often ranging from "not very happy" to "very happy." In either case, the outcome is an ordinal variable, which ranks the alternative

[8] The topic has attracted considerable attention within the literature on the "capability approach." See, among others, Sen (1985, 1992), Alkire (2002, 2007), Nussbaum (1990, 1993, 2003), Kuklys (2005), Robeyns (2005, 2006), and Basu and López-Calva (2011).

[9] Well-known examples are the Programme for International Student Assessment (PISA) for 15-year-old students and the Programme for the International Assessment of Adult Competencies (PIAAC), both coordinated by the Organisation for Economic Co-operation and Development (OECD). Micklewright and Schnepf (2007, p. 133) compare the cross-country inequality in learning achievement scores and call for caution in the use of the income inequality measurement toolbox, because "it is doubtful whether the measurement of the scores is on a ratio scale. Their nature is therefore quite different from that of data on income or height."

ratings without however providing any information on how much one rating is better, or worse, than another rating.

Cardinal continuous variables, such as income, probably represent a minority of available indicators of well-being. The application of measurement tools that are standard in income distribution analysis may hence need to be reconsidered if applied to nonmonetary domains.[10] This warning applies to, but is clearly not exclusive of, multidimensional analysis. One specific problem that arises in this context concerns the commensurability of the indicators when they are merged into a single index. It is generally tackled by employing procedures of standardization that, for instance, transform the original variable by taking its (normalized) distance from benchmark values (for some examples of these transformations, see Decancq and Lugo, 2013, p. 12). Alternatively, ordinal criteria might also be applied to quantitative variables (e.g., by classifying units according to the quantile to which they belong).[11] Irrespective of the specific procedure adopted, the transformation of the original values substantially affects the outcome.

Many variables are dichotomous, or binary, either by definition or after a comparison of individual achievement with some social norm: for instance, we may classify those deprived in housing conditions as all individuals living in households with less than one room per person, transforming the variable "room per person" into a binary one. The use of dichotomous variables is at the center of the counting approach examined below.

In poverty assessments, the choice of the indicators is intertwined with the definition of the respective deprivation thresholds. This problem parallels the problem for income or consumption in univariate analysis, with absolute, relative, subjective, and legal criteria being the main alternatives (e.g., Callan and Nolan, 1991). In multivariate analyses, these problems may be amplified by the consideration of intangible dimensions for which it is more contentious to identify minimum thresholds (Thorbecke, 2007). Similar to the univariate case, however, the binary distinction between a "bad state" and a "good state" may be too sharp because deprivation might occur by degrees. Moving along these lines, Desai and Shah (1988) focus on the distance of the individual achievements from modal values in each dimension, taken to represent the social norm, whereas the extensive

[10] Researchers are increasingly exploring the measurement of inequality when using qualitative ordinal variables, such as self-reported health status (e.g., Abul Naga and Yalcin, 2008; Allison and Foster, 2004; Bleichrodt and van Doorslaer, 2006; van Doorslaer and Jones, 2003) and happiness (e.g. Dutta and Foster, 2013; Kalmijn and Veenhoven, 2005). Cowell and Flachaire (2012) have axiomatically developed a class of inequality indices for categorical data, conditional on a reference point, which are based on the individuals' positions in the distribution. Zheng (2008) suggests that, when data are ordinal, stochastic dominance has limited applicability in ranking social welfare and no applicability in ranking inequality.

[11] Qizilbash (2004) discusses the sensitivity of empirical estimates of South African poverty to the transformation of indicators from cardinal to ordinal, using the Borda score, as well as to variation in the thresholds used to define deprivation.

literature on the "fuzzy sets approach" formalizes a continuum of grades of poverty by means of a "membership" function.[12] Such a membership function may assume any value between 0 and 1: the two extreme values indicate that a person is definitely nondeprived (0) or deprived (1), and all other values indicate partial membership in the pool of the deprived. The form of the membership function plays a crucial role in the construction of a fuzzy deprivation measure. Although largely seen as a distinct approach in the multivariate analysis of deprivation, there is nothing inherently multidimensional in the theory of fuzzy sets.

3.2.3 Weighting of Dimensions

Weights determine the extent to which the selected attributes contribute to well-being and the degree by which we can substitute one attribute for another, interacting with the functional form used to aggregate dimensions. This can be easily seen by defining individual well-being S_β as the weighted mean of order β of the achievements in the r dimensions, as suggested, for instance, by Maasoumi (1986),

$$S_\beta = \begin{cases} \left[\sum_{k=1}^{r} w_k x_k^\beta\right]^{1/\beta} & \beta \neq 0 \\ \prod_{k=1}^{r} x_k^{w_k} & \beta = 0 \end{cases} \tag{3.1}$$

where x_k is nonnegative and represents the level of attribute k, $k=1,2,\ldots,r$, and w_k is the corresponding weight. Notice that expression (3.1) turns into an index of deprivation if the r attributes measure hardship. The weights w_k and the parameter β jointly govern the degree of substitution between any pair of cardinal attributes. Indeed, the marginal rate of substitution between attributes b and a, which is the quantity of b that has to be given up in exchange for one more unit of a in order to leave well-being unchanged, is equal to:

$$MRS_{b,a} = \frac{\mathrm{d}x_{bi}}{\mathrm{d}x_{ai}} = -\left(\frac{w_a}{w_b}\right)\left(\frac{x_a}{x_b}\right)^{\beta-1}. \tag{3.2}$$

If $\beta=1$ well-being is simply the (weighted) arithmetic mean of the achievements in all dimensions, which are then perfectly substitutable at a rate equal to the ratio of their

[12] See Cerioli and Zani (1990), Cheli et al. (1994), Cheli (1995), Cheli and Lemmi (1995), Chiappero Martinetti (1994, 2000), Betti et al. (2002), Dagum and Costa (2004), Qizilbash and Clark (2005), Betti and Verma (2008), Betti et al. (2008), Belhadj (2012), and Belhadj and Limam (2012). Deutsch and Silber (2005), Pérez-Mayo (2007), and D'Ambrosio et al. (2011) compare empirical results for multidimensional measures of poverty based on the fuzzy sets approach with those derived from applying alternative approaches (axiomatic approach, information theory, efficiency analysis, latent class analysis). Kim (2014) studies the statistical behavior of fuzzy measures of poverty.

respective weights. In all other cases, the marginal rate of substitution also depends on relative achievements: the further away β is from 1, the more an unbalanced achievement in the two dimensions matters. When β goes to infinity (minus infinity), the attributes are perfect complements, and the well-being level depends on the highest (lowest) achievement, regardless of the values assigned to the weights.

The pattern of substitution among attributes can be more muddled than in (3.2) when the functional form of the well-being aggregator is more complex than (3.1), but it is bound to depend critically on weights, except in the extreme cases in which the attributes are perfect complements. The choice of weights might have a significant effect on the results of multidimensional analyses of inequality and poverty. For instance, Decancq et al. (2013) find that the identification of the worst-off in a sample of Flemish people is considerably influenced by the use of alternative weighting schemes of the attributes. In a comparison of the incidence of income-and-health poverty in selected European countries from 2000 to 2001, Brandolini (2009) finds that the ranking of Italy and Germany reverses as weights are shifted from one dimension to the other, although the ordering of France and the United Kingdom mostly remains unchanged. Here, we outline approaches to weighting by drawing on Brandolini and D'Alessio (1998), and we refer to Decancq and Lugo (2013) for a more comprehensive discussion.

A popular way of setting weights is to treat all attributes equally. This is the case with the Human Development Index, which assigns the same weight (one-third) to the three basic dimensions considered: a long and healthy life, access to knowledge, and a decent standard of living (e.g., UNDP, 2013). Equal weighting may result from either an "agnostic" attitude and a wish to reduce interference to a minimum or from the lack of information about some kind of "consensus" view. For instance, Mayer and Jencks (1989, p. 96) opt for equal weighting, after remarking that "ideally, we would have liked to weight [the] ten hardships according to their relative importance in the eyes of legislators and the general public, but we have no reliable basis for doing this." (In fact, there may be disagreement among the legislators and the public, let alone within the public itself.)

Some departure from equal weighting is envisaged by Atkinson et al. (2002) and Marlier and Atkinson (2010). They propose a set of principles for the design of social indicators for policy purposes, among which is the principle that the weights should be "proportionate," so that dimensions have "degrees of importance that, while not necessarily exactly equal, are not grossly different" (Marlier and Atkinson, 2010, p. 289). This criterion only sets some reasonable boundaries, without specifying how to define unequal weights.

The social evaluator can directly elicit the weighting structure from consultations with groups of experts or the public at large or from the importance assigned to dimensions of well-being by survey respondents, or the evaluator can indirectly generate the structure from estimates of happiness equations.[13] The last procedure is followed by Decancq et al. (2014) who characterize axiomatically a class of multidimensional poverty indices that are consistent with individual preferences in the aggregation of the different dimensions. In addition to standard axioms, they postulate principles for interpersonal poverty comparisons that lead to measuring individual poverty as a function of the fraction of the poverty line vector to which the agent is indifferent. The poverty threshold is therefore defined in terms of well-being using person-specific weights. In some exercises, users of statistics are allowed to build their own sets of weights. For instance, the OECD Better Life Index allows people to compare well-being across countries by means of eleven indicators of quality of life that can be rated equally or according to individual preferences (see Boarini and Mira D'Ercole, 2013 and the initiative's website http://www.oecdbetterlifeindex.org/). In all these cases, the choice of weights relies on some implicit or explicit normative criterion.

Under certain hypotheses, market prices provide weights that capture a trade-off between dimensions that is consistent with consumer welfare. Sugden (1993) and Srinivasan (1994) contend that the availability of such an "operational metric for weighting commodities" makes traditional real-income comparison superior in practice to Sen's capability approach. Ravallion (2011a, p. 243) argues that the main multidimensional poverty indices aggregate deprivations in a manner that "essentially ignores all implications for welfare measurement of consumer choice in a market economy. While those implications need not be decisive in welfare measurement, it is clearly worrying if the implicit tradeoff between any two market goods built into a poverty measure differs markedly from the tradeoff facing someone at the poverty line." On the other hand, market prices may be distorted by market imperfections and externalities, and they do not exist for many constituents of well-being and their imputation may be arduous, although various approaches estimate the "willingness to pay" in order to add the monetary value of nonincome dimensions to income (e.g. Becker et al., 2005; Fleurbaey and Gaulier, 2009; see Chapter 2). More importantly, they may be conceptually inappropriate for welfare comparisons, a task for which they are not devised (Foster and Sen, 1997; Thorbecke, 2007).

The main alternative and widely applied approach is "to let the data speak for themselves." Methods differ, but we may cluster them into two main categories: frequency-based approaches and multivariate statistical techniques. Since Desai and Shah (1988) and Cerioli and Zani (1990), many researchers assume that the smaller the proportion of people with a certain deprivation, the higher the weight that should be assigned to that deprivation, on the grounds that a hardship shared by few is more important than one shared by many. This approach raises two problems. First, it may lead

[13] See Decancq and Lugo (2013, pp. 24–26) for a discussion, and Bellani (2013), Bellani et al. (2013), Cavapozzi et al. (2013), Decancq et al. (2013), and Mitra et al. (2013) for some examples.

to a questionably unbalanced structure of weights. As observed by Brandolini and D'Alessio (1998), in 1995, the shares of Italians with low achievement in health and education were 19.5% and 8.6%, respectively. With these proportions, education insufficiency would be valued more than health insufficiency: one-tenth more according to Desai and Shah's formula, and over one-half more according to Cerioli and Zani's formula. Whether education should attain a weight so much higher than health is certainly a matter of disagreement. This criterion also makes the weights endogenous to the distributions being studied. Thus, it implies that we should take country-specific weights in an international comparison of multidimensional poverty, unless we impose a common, but arbitrary, set of weights. This observation also applies to the suggestions by Betti et al. (2008), who suggest taking weights proportional to the dispersion of the attributes in the population (adjusted for their bilateral correlations to avoid redundancy), and by Vélez and Robles (2008), who select weights that allow a set of multidimensional poverty measures to better track the dynamics of self-perceived well-being.

Several multivariate statistical techniques are employed to aggregate dimensions.[14] Maasoumi and Nickelsburg (1988), Klasen (2000), and Lelli (2005) use the analysis of principal components, on the grounds that this approach "... uncovers empirically the commonalities between the individual components and bases the weights of these on the strength of the empirical relation between the deprivation measure and the individual capabilities" (Klasen, 2000, p. 39, fn. 13). Schokkaert and Van Ootegem (1990), Nolan and Whelan (1996a,b), and Whelan et al. (2001) use factor analysis to aggregate elementary indicators into measures of well-being or deprivation. These papers tend to use this technique to identify few distinct constituents of well-being, however: as noted by Schokkaert and Van Ootegem (1990, p. 439–40), their application of factor analysis is "a mere data reduction technique," which does not provide any indication about the relative valuation of each attribute. Several authors apply latent variable models or structural equation modeling to collapse multiple indicators into indices of total or domain-specific deprivation (Ayala et al., 2011; Di Tommaso, 2007; Krishnakumar, 2008; Krishnakumar and Ballon, 2008; Krishnakumar and Nagar, 2008; Kuklys, 2005; Navarro and Ayala, 2008; Pérez-Mayo, 2005, 2007; Tomlinson et al., 2008; Wagle, 2005, 2008a,b). Dewilde (2004) uses a two-step latent class analysis, evaluating deprivation in specific domains in the first step and the latent concept of overall poverty in the second step. Lovell et al. (1994), Deutsch and Silber (2005), Ramos and Silber (2005), Anderson et al. (2008), Ramos (2008), and Jurado and Pérez-Mayo (2012) apply methods developed in efficiency

[14] For a discussion of applied multivariate techniques, see Sharma (1996). Moreover, see Ferro Luzzi et al. (2008), Pisati et al. (2010), Whelan et al. (2010), Lucchini and Assi (2013), and Caruso et al. (2014) for an application of cluster analysis to identify population subgroups homogeneous by well-being or deprivation level, and Hirschberg et al. (1991) for an analogous comparison across countries, as well as Asselin and Anh (2008) and Coromaldi and Zoli (2012) for an application of multiple correspondence analysis and nonlinear principal component analysis, respectively.

analysis to aggregate the various attributes of well-being. These methods allow estimating the level of individual achievement relative to the achievement frontier, providing implicit estimates of the values of the weights. In a related approach, Cherchye et al. (2004) construct a synthetic indicator to assess European countries' performance in achieving social inclusion, with weights being variable in order to provide the most favorable evaluation for each country. They contend that this approach preserves the "legitimate diversity" of countries in pursuing their own policy objectives, because a relatively better performance in a particular dimension is seen as revealing a policy priority.

The methods reviewed in the next sections generally allow for the possibility that weights can differ across dimensions in the social evaluation of poverty and inequality. Our brief overview suggests some ways to define them. Two comments are in order. First, multivariate statistical techniques differ from other approaches in that their aim is to estimate the level of individual achievement; weights are integral part of the aggregation procedure and have no truly independent meaning. We may then wonder whether it is appropriate to use them in conjunction with many of the methods discussed below. Second, as the weighting structure captures the importance assigned to each attribute, it is bound to reflect different views. On one side, this suggests questioning the use of techniques that may be robust from a statistical viewpoint but ignore the intrinsically normative aspect of the choice of weights. On the other side, it hints that one way to account for this plurality of views is to specify ranges of weights rather than a single set of weights, although this approach might lead to a partial ordering, as suggested by Sen (1987, p. 30; see also Foster and Sen, 1997, p. 205).[15]

3.3. MULTIDIMENSIONAL POVERTY MEASUREMENT

A long tradition in social sciences has been concerned with measuring material deprivation by looking at a number of indicators of living conditions, such as the ownership of durables or the possibility of carrying out certain activities such as going out for a meal with friends. The typical way to summarize the information has been to count the number of dimensions in which people fail to achieve a minimum standard, and, hence, this method is labeled the "counting approach." It represents the simplest way to embed the association between deprivations at the individual level into an overall index of deprivation.

In the counting approach, the synthesis of the available information begins with aggregating across the single dimensions for each individual, and then across the individuals themselves. However, we could invert the order of aggregation by computing the proportions of people suffering in each dimension and then aggregating these proportions into a composite index of deprivation. This different order of aggregation has the great advantage of allowing us to draw these proportions from various sources. This

[15] Cherchye et al. (2008) present a methodology that incorporates a range of weighting schemes in the ranking of vectors of attributes.

characteristic makes this "composite index" approach easily understandable and very popular, especially in public debates in which there is a need to summarize headline messages from sets of indicators. If the dimensions of well-being are independent of each other, the order of aggregation does not matter, and the two approaches are equivalent. However, if they are dependent and suffering from multiple deprivations has a more than proportionate effect on people's well-being, ignoring the impact of the association among the achievements in the various dimensions, as with the composite index approach, may imply overlooking an important aspect of hardship. This is not the case for an indicator such as severe material deprivation in the Europe 2020 strategy, because it would rank a society with one person suffering from four deprivations and three persons not suffering from any differently from a society in which four people fail in one dimension each.

The relationship between the two approaches can be better understood by considering the simple situation in which only two dimensions exist. Assume that X_k is equal to 1 if an individual suffers from deprivation in dimension k and 0 otherwise, with $k=1,2$. Let $p_{ij}=\Pr((X_1=i)\cap(X_2=j))$, $p_{i+}=\Pr(X_1=i)$, and $p_{+j}=\Pr(X_2=j)$. Then, assign equal weight to the two deprivation indicators and define the deprivation score $X=X_1+X_2$, which can take the values $(0,1,2)$ with associated probabilities (q_0,q_1,q_2). The parameters (q_0,q_1,q_2) of the count distribution X are determined by the parameters of the original two-dimensional simultaneous distribution in the following manner: $q_0=p_{00}$, $q_1=p_{10}+p_{01}$, and $q_2=p_{11}$. The original and derived distributions are summarized in Table 3.1.

If only the marginal distributions in the left panel of Table 3.1 were known, an overall poverty indicator P could be expressed as a function g of p_{1+} and p_{+1} only, that is $P=g(p_{1+},p_{+1})$, which is an example of a composite poverty index. If the simultaneous distribution was known, we could turn to the distribution of X in the right panel of Table 3.1, and the overall index could account for the number of deprivations that each individual suffers from. Counting deprivations highlights two possible ways of identifying someone as poor: either he fails in either single dimension ($X=1$), or he fails in both ($X=2$). In the first case, we adopt the union criterion: the poor are those with at least one deprivation and $P=g(1-p_{00})$. In the second case, we favor the intersection criterion: the poor are those with two failures and $P=g(p_{11})$. The contrast between union and

Table 3.1 The distribution of deprivations in two dimensions and the derived distribution of deprivations scores

	$X_2=0$	$X_2=1$	
$X_1=0$	p_{00}	p_{01}	p_{0+}
$X_1=1$	p_{10}	p_{11}	p_{1+}
	p_{+0}	p_{+1}	1

	$X=X_1+X_2$
$X=0$	$q_0=p_{00}$
$X=1$	$q_1=p_{10}+p_{01}$
$X=2$	$q_2=p_{11}$
	1

Source: authors' elaboration.

intersection criteria plays a fundamental role in the measurement of multidimensional deprivation (see Atkinson, 2003). It also suggests that the occurrence of deprivation in some domains need not entail a condition of overall poverty: if we adopt the intersection criterion, only those with two failures are regarded as poor individuals, whereas those with only one failure are not. Setting a critical number of dimensions c, $1 \leq c \leq r$, to identify the poverty status introduces an additional threshold over those already set for defining deprivation in each dimension (see Alkire and Foster, 2011a,b). We return to this issue in Section 3.3.2.6.

The available information may be richer than the knowledge about the deprived/not deprived status in a number of dimensions, however. Rather than dichotomous, variables may be continuous or discrete with at least three categories. We may then want the overall poverty indicator to account not only for the occurrence of deprivation, that is, an individual achievement below the given dimension-specific threshold, but also for its intensity, that is, the shortfall of this achievement as compared to the threshold.

These observations illustrate that the reach of the informational basis conditions the multidimensional methods that can be used to measure poverty. When individual-level data on multiple attributes are not available, a composite index may be the only measure that can be calculated. When these data exist but are not publicly available, multidimensional poverty analysis may still be possible by using counting measures, if statistical offices release simple tabulations such as those discussed in the examples in Section 3.3.2. We use the complexity of informational needs as the criterion to organize the discussion of this section. We begin with the composite multidimensional poverty indices that only require information on the marginal distributions and can be estimated by gathering data from separate sources. All other multidimensional measures need an integrated database in which the information for each relevant dimension is available for each individual unit. We first consider counting measures that use minimal information: the distribution of the population by number of deprivations. With r dimensions, it is sufficient to know r values (the proportions of the population suffering from deprivation in $0, 1, \ldots, r$ dimensions). Although it is the oldest multidimensional approach in social sciences, the counting approach is arguably the least structured from a theoretical point of view, and we devote relatively more space to its examination. Due to its simplicity, the counting approach offers transparent illustrations of alternative aggregation methods, as well as the role of various normative rearrangement principles, and it helps to clarify the distinction between deprivation and poverty. Next, we turn to multidimensional poverty indices requiring the knowledge of individual achievements in each dimension. Lastly, we discuss criteria for partial ordering.

3.3.1 The Composite Index Approach

We can measure the overall poverty of a society by aggregating over the proportions of individuals suffering from deprivation in the r dimensions of well-being, whenever this is the only available information. A prominent example of this composite index approach is the Human Poverty Index (*HPI*), which was published by the United Nations Development Program from 1997 to 2009 (UNDP, 1997). As originally formalized by Anand and Sen (1997), a general version of the index with r dimensions, weighted by w_k, is defined by

$$HPI_\beta = \zeta_1(p_1, p_2, \ldots, p_r) = \left(\sum_{k=1}^{r} w_k p_k^\beta\right)^{1/\beta}, \tag{3.3}$$

where p_k is the proportion suffering from deprivation in dimension k (in the two-dimensional case of Table 3.1 $p_1 = p_{1+}$ and $p_2 = p_{+1}$), $\beta > 0$, and $w_k > 0$ for all k; if the r dimensions are equally weighted, $w_k = 1/r$. As β rises, greater weight is given to the dimension in which there is the most deprivation. UNDP (1997) paid particular attention to three dimensions related to longevity, knowledge, and a decent standard of living, and it later added a fourth dimension, social exclusion, for rich countries. In either case, β was set equal to 3 to give "additional but not overwhelming weight to areas of more acute deprivation" (UNDP, 2005, p. 342).[16]

Bossert et al. (2013) provide an axiomatic characterization of (3.3) for the case in which $\beta = 1$, based on the condition of additive decomposability in attributes as well as in individuals (see also Pattanaik et al., 2011). This case is of some interest: it assumes perfect substitutability among the components, and the index HPI_1 equals the weighted arithmetic mean of the headcount indices across all dimensions. This implies that people who suffer from k deprivations, with $0 \leq k \leq r$, are counted k times by the index HPI_1. Although rather crude and *ad hoc*, this is a simple way of giving heavier weight to people suffering from multiple deprivations. The implicit assumption is that the effect of deprivations is proportionate, however: suffering from two deprivations is twice as bad as suffering from one. If there are reasons to question this assumption, then the inability of *HPI*-type measures to discriminate between situations in which deprivations are concentrated on few people and situations where an identical total amount of deprivations is spread across many people represents a serious shortcoming.

Dutta et al. (2003) prove that composite indices can lead to the same conclusions as those that would be derived from aggregating first across dimensions and then across individuals only under very restrictive conditions on the aggregation functions. Namely, "the overall deprivation of an individual must be a weighted average of her deprivations [i.e., proportionate shortfalls relative to benchmark values] in terms of the different attributes, and society's overall deprivation must be a simple average of the overall deprivation levels of the different individuals in the society" (Dutta et al., 2003, p. 202). Both conditions may be debatable: the first because it implies that marginal rates of substitution between any pair of attributes are insensitive to the depths of deprivations; the second because it is liable to the same criticism leveled against the poverty gap by Sen (1976). Analogous

[16] Chakravarty and Majumder (2005) characterize a general family of deprivation indices that includes an index ordinally equivalent to *HPI* as a member.

results hold when the equivalence condition is set with respect to rankings rather than indices. Pattanaik et al. (2011) discuss further weaknesses of *HPI*-type measures.

Although composite indices may not be consistent with an approach that sees society's overall poverty as a function of individual poverty levels, as happens in standard welfare economics, they might be justified by taking a different set of ethical assumptions.

3.3.2 The Counting Approach

In many cases, we know more than the headcount poverty ratio for each dimension, and we observe how many people are suffering from deprivation in one dimension, two dimensions, and so forth. Counting the number of failures is well rooted in the analysis of deprivation in social sciences, but the characteristics of the underlying social judgments and the relationship with standard welfare approaches still need clarification. Atkinson (2003), for instance, draws a parallel between the difficulty of deriving dominance conditions in the counting case and the failure of the headcount poverty measure to satisfy the Pigou–Dalton principle of transfers in the one-dimensional case. However, this difficulty stems from defining welfare criteria in terms of the distributions of the underlying continuous variables across people rather than in terms of the distribution of deprivation scores. As the deprivation score counts the number of dimensions in which an individual fails to achieve the minimum standards, it is by definition a discrete variable ranging from 0 to the number of dimensions considered. The distribution of deprivation scores contains all the relevant information in the counting approach, which by construction implies neglecting levels of achievement in the original variables. Dominance conditions in the counting approach can be established following this line of reasoning. In this section, we discuss these conditions, and we show how they can yield counting measures that encompass those proposed by Atkinson (2003), Chakravarty and D'Ambrosio (2006), and Alkire and Foster (2011a,b).

As is standard in the counting literature, we assume that individuals might suffer from deprivation in r different dimensions, and then we sum the number of actual deprivations.[17] Let X_i be equal to 1 if an individual suffers from deprivation in the dimension i and 0 otherwise. Moreover, let

$$X = \sum_{i=1}^{r} X_i$$

[17] Cappellari and Jenkins (2007) observe that the practice of constructing raw deprivation sum-scores is "ubiquitous" but has weak theoretical foundations. They suggest that a promising alternative way to summarize multiple deprivations can rely on the item response modeling approach used in psychometrics and educational testing, although they find similar results in a comparison of the two approaches for British data.

be a random discrete variable with cumulative distribution function F and mean μ, and let F^{-1} denote the left inverse of F. Thus, $X=1$ means that the individual suffers from one deprivation, $X=2$ means that the individual suffers from two deprivations, and so on. We call X the deprivation count and F the deprivation count distribution. Furthermore, let $q_k = \Pr(X=k)$, which yields

$$F(k) = \sum_{j=0}^{k} q_j, \quad k = 0, 1, \ldots, r \tag{3.4}$$

and

$$\mu = \sum_{k=1}^{r} k q_k. \tag{3.5}$$

For the sake of simplicity, we are assigning equal weight to all dimensions, but this assumption can be relaxed (see Section 3.3.2.5).

In order to compare count distributions, we introduce appropriate dominance criteria to obtain partial orderings (Section 3.3.2.1) and complete orderings (Sections 3.3.2.2–3.3.2.4).[18] Although the multidimensional approaches discussed in Section 3.3.3 focus on the distribution of people's achievements, the dominance criteria formulated for the counting approach are defined in terms of the distribution F of the univariate discrete variable X.

3.3.2.1 Partial Orderings

As is standard in the income distribution literature, the first criterion regards first-degree dominance.[19]

Definition 3.1

A deprivation count distribution F_1 is said to first-degree dominate a deprivation count distribution F_2 if

$$F_1(k) \geq F_2(k) \quad \text{for all } k = 0, 1, \ldots, r$$

and the inequality holds strictly for some k.

If F_1 first-degree dominates F_2, then F_1 exhibits less deprivation than F_2. An example is given in Figure 3.1, in which we use the material deprivation indicators in five European countries in 2012 drawn from Eurostat (2014) and reported in Table 3.2. Figure 3.1 plots on the vertical axis the cumulative proportion of persons who suffer from

[18] Lasso de la Vega (2010) and Yalonetzky (2014) also identify dominance conditions to rank deprivation count distributions.

[19] The first-degree stochastic dominance relations for integer variables representing the counting of people achievements, rather than deprivations, are studied by Chakravarty and Zoli (2012).

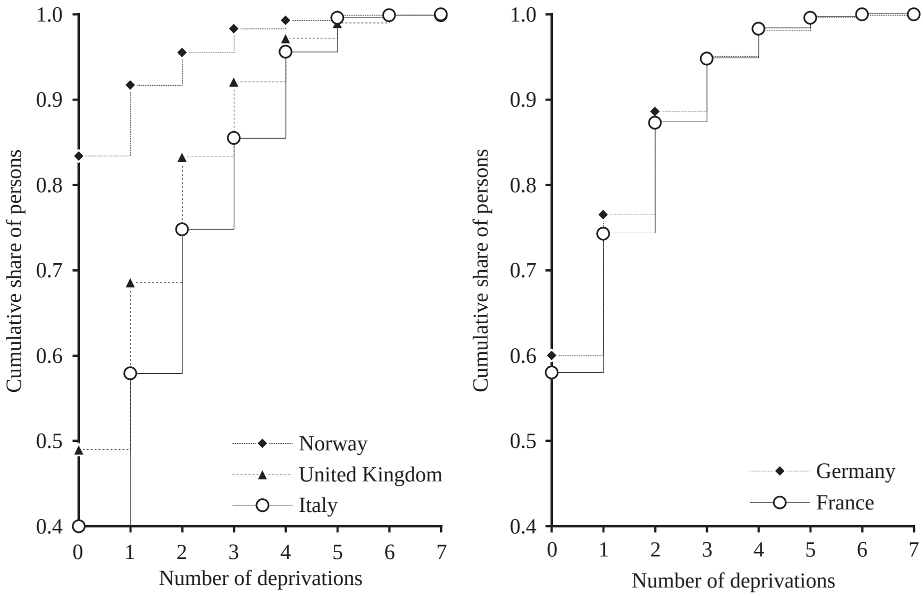


Figure 3.1 Cumulative distributions of material deprivation scores in selected European countries in 2012. *Source: authors' elaboration on data from Eurostat (2014).*

Table 3.2 Distribution of material deprivations in selected European countries in 2012 (percentage of total population)

Number of deprivations	France	Germany	Italy	Norway	United Kingdom
None	58.0	60.0	39.6	83.4	49.0
1 item	16.3	16.5	18.3	8.3	19.6
2 items	13.0	12.1	16.9	3.8	14.7
3 items	7.5	6.5	10.7	2.8	8.8
4 items	3.5	3.0	10.1	1.0	5.1
5 items	1.3	1.5	4.0	0.6	1.8
6 items	0.4	0.3	0.3	0.0	0.9
7 items	0.0	0.1	0.1	0.1	0.1
8 items	0.0	0.0	0.0	0.0	0.0
9 items	0.0	0.0	0.0	0.0	0.0
All	100.0	100.0	100.0	100.0	100.0

Source: Eurostat (2014).

deprivation in, at most, the number of dimensions indicated on the horizontal axis. (Figure 3.1 considers a maximum of seven deprivation items because nobody suffers from more than seven in the countries considered.) The left panel shows that Norway first-degree dominates both the United Kingdom and Italy, whereas the last two countries cannot be ordered by the criterion of first-degree dominance because their distributions intersect. The United Kingdom clearly lies ahead of Italy for up to five items, but then it exhibits a share of people suffering from six or seven deprivations that is more than twice the Italian level (1% vs. 0.4%, see Table 3.2). The right panel of Figure 3.1 shows that the cumulative distributions of deprivation scores for France and Germany also intersect, though they are much closer. The share of nondeprived is higher in Germany than in France, and the same holds true when we sequentially add those with one, two, and three deprivations; however, when we add people suffering from four deprivations, the order reverses, and it no longer changes when we consider more severe situations.[20]

This example shows that first-degree dominance might be too demanding in practice: where count distributions intersect, they can be ranked only by defining weaker dominance criteria. This implies that we have to impose stricter conditions on the preference ordering of the social evaluator, taking into account that, in the study of deprivation, we might be leaning toward either the intersection or the union criteria. In the former case, we would start aggregating "from above," looking first at the proportion of those who are deprived in r dimensions, then adding the proportion of those failing in $r-1$ dimensions, and so forth; in the latter case, we would start "from below." This distinction naturally leads to the definition of two second-degree dominance criteria, as suggested by Aaberge and Peluso (2011):

Definition 3.2A

A deprivation count distribution F_1 is said to second-degree downward dominate a deprivation count distribution F_2 if

$$\sum_{k=s}^{r} F_1(k) \geq \sum_{k=s}^{r} F_2(k) \quad \text{for all } s = 0, 1, \ldots, r$$

and the inequality holds strictly for some s.

Definition 3.2B

A deprivation count distribution F_1 is said to second-degree upward dominate a deprivation count distribution F_2 if

$$\sum_{k=0}^{s} F_1(k) \geq \sum_{k=0}^{s} F_2(k) \quad \text{for all } s = 0, 1, \ldots, r$$

and the inequality holds strictly for some s.

[20] In this example and in all subsequent empirical illustrations, we treat statistics as if they were exact, and we abstract from the fact that they are subject to sampling and other types of errors. Accounting for these errors would possibly lead us to conclude that neither the observed difference between France and Germany nor the upper tail intersection between France and Norway is statistically significant.

If F_1 second-degree dominates F_2, then F_1 exhibits less deprivation than F_2, as before, but this result is now obtained at the cost of imposing the stricter conditions on the preference ordering that will be shown below by Theorems 3.1A and 3.1B. Moreover, we have to make a choice between being more concerned with the extent to which deprivation is diffused across the population (union criterion) or the occurrence of multiple deprivations (intersection criterion). In the first case, we would adopt second-degree upward dominance. Intuitively, we can see this in Definition 3.2B from the fact that we are making comparisons on (doubly) cumulated population proportions that start by considering the share of people who do not suffer from any deprivation, $F(0)$, and we sequentially add the shares of those who suffer from one deprivation, then those who suffer from two deprivations, and so forth. In calculating the cumulative function we "go up." The opposite happens in the second case, for which we aggregate "going down," thus placing more weight on the most deprived. Formally, second-degree upward dominance parallels the dominance criterion used by Atkinson (1970) for ranking income distributions. Second-degree downward dominance has no correspondent in the income inequality literature, because it would be inconsistent with the Pigou–Dalton principle of transfers. It is, however, analogous to the criterion introduced for Lorenz curves by Aaberge (2009).

Is agreeing on whether to go up (union criterion) or to go down (intersection criterion) when we aggregate deprivation scores sufficient in empirical applications? Not always. This can be seen by reconsidering the previous comparisons of Italy and the United Kingdom, and of France and Germany, with neither country in each comparison found to first-degree dominate the other. In Figure 3.2, we plot the difference between the integrated cumulative distributions considered by Definitions 3.2A and 3.2B for each pair of countries. If we integrate going up as in Definition 3.2B, the United Kingdom and Germany second-degree (upward) dominate Italy and France, respectively: the lower proportions of people who do not suffer from any deprivation give the first two countries an advantage that is not offset by their worst results for the incidence of people deprived in many dimensions. On the other hand, if we integrate going down as in Definition 3.2A, the difference between the integrated cumulative distributions changes from positive to negative, and no country second-degree (downward) dominates the other in either comparison. The distribution of deprivation scores enables social evaluators favoring the union perspective to rank the United Kingdom and Germany ahead of Italy and France, but it does not allow social evaluators supporting the intersection perspective to draw unambiguous conclusions. In such a case, higher-degree criteria are needed, although they could still provide a partial ordering. The exploration of higher-order dominance criteria is a topic for further research. We turn instead to methods that can lead to a complete ordering.

3.3.2.2 Complete Orderings: The Independence Axioms

A complete ordering can be achieved by imposing an independence axiom for preference ordering. This allows us to weight differently certain parts of the distributions and eventually to define a summary measure of deprivation. Formally, let social preferences be

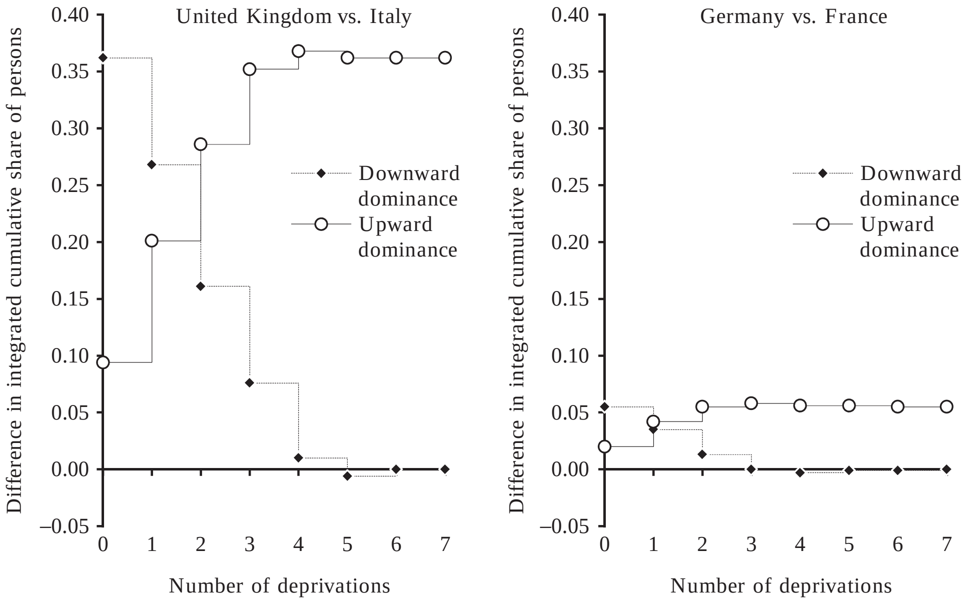


Figure 3.2 Second-degree dominance for material deprivation scores in selected European countries in 2012. *Source: authors' elaboration on data from Eurostat (2014).*

represented by the ordering $\succeq$ defined on the family of deprivation count distributions F. This preference ordering is assumed to be continuous, transitive, and complete and to satisfy the condition of first-degree count distribution dominance. As proved by Debreu (1964), a preference ordering that is continuous, transitive, and complete can be represented by a continuous and increasing preference functional. We need further conditions to give social preferences an explicit empirical content, however. We therefore introduce two alternative independence conditions, which require that the preference ordering is invariant with respect to certain changes in the count distributions being compared:

Axiom (Independence). *Let F_1 and F_2 be members of $\boldsymbol{F}$. Then $F_1 \succeq F_2$ implies $\alpha F_1 + (1-\alpha)F_3 \succeq \alpha F_2 + (1-\alpha)F_3$ for all $F_3 \in \boldsymbol{F}$ and $\alpha \in [0,1]$.*

This axiom focuses on the proportions of people suffering from given numbers of deprivations (the F). We could instead focus on the number of deprivations that is associated with a given proportion of people, that is, more technically, the rank in the count distribution (the F^{-1}). This corresponds to an alternative version of the independence axiom, as in the literatures on uncertainty and inequality:

Axiom (Dual Independence). *Let F_1 and F_2 be members of $\boldsymbol{F}$. Then $F_1 \succeq F_2$ implies $\left(\alpha F_1^{-1} + (1-\alpha)F_3^{-1}\right)^{-1} \succeq \left(\alpha F_2^{-1} + (1-\alpha)F_3^{-1}\right)^{-1}$ for all $F_3 \in \boldsymbol{F}$ and $\alpha \in [0,1]$.*

If F_1 is weakly preferred to F_2, then the independence axiom (similar to the expected utility theory) states that any mixture on F_1 is weakly preferred to the corresponding mixture on F_2: identical mixing interventions on the count distributions do not affect their ranking, which depends solely on how the differences between the mixed count distributions are judged. Thus, if the overall count deprivation is lower in country 1 than in country 2, so that $F_1 \succeq F_2$, the ranking would not change by adding to the population of either country the same group of migrants, whose deprivation distribution is F_3. The ordering relation $\succeq$ is therefore invariant with respect to the aggregation of subpopulations across deprivations.

The dual independence axiom shifts toward aggregating subsets of deprivation dimensions across proportions of people. Assume that there are only two deprivation indicators, income and health, and that two alternative tax and benefit regimes produce the two count deprivation distributions F_1 and F_2 for income. Next, match F_1 with the count deprivation distribution F_3 for health in such a way that the most deprived person in income is also the most deprived person in health, the second most deprived person in income is the second most deprived person in health, and so on. Match F_2 and F_3 in the same way. If the count deprivation distribution F_1 is preferred to F_2 for income, then the share of income-deprived people under regime 1 is lower than the corresponding share under regime 2. Dual independence means that, given any distribution F_3 of health deprivation counts, F_1 will continue to be preferred to F_2 after matching both F_1 and F_2 with F_3.[21] The dual independence axiom imposes this invariance property regardless of the shape of the count deprivation distribution for health (F_3) and of the weights used for such a matching (α).

The essential difference between the two axioms is that the independence axiom deals with the relationship between a given number of deprivations and weighted averages of the corresponding population proportions, but the dual independence axiom deals with the relationship between given population proportions and weighted averages of the corresponding numbers of deprivations. No one has so far provided a convincing justification for preferring one axiom to the other, but the choice of the axiom yields summary measures of deprivation with different decomposition properties. For instance, indices consistent with the independence axiom can be expressed as weighted averages of the corresponding indices computed for mutually exclusive population subgroups, whereas the indices satisfying the dual independence axiom cannot. By contrast, the dual measures offer a convenient decomposition by sources of deprivation, whereas the measures associated with the independence axiom cannot. Moreover, as measures of income

[21] This argument parallels the rationale offered by Weymark (1981, p. 418) for his "Weak Independence of Income Source" axiom: "if in two income distributions the incomes from all but one type of income are the same in both distributions, then the overall judgement that one distribution is more unequal than a second is completely determined by a comparison of the distributions of income from the variable source." Gajdos and Weymark (2005) call the corresponding multidimensional condition "Weak Comonotonic Additivity."

inequality, they have the convenient property of being expressed as linear functionals of the Lorenz curve, whereas the primal measures cannot.

The "primal approach," based on the independence axiom, is analogous to the inequality framework developed by Atkinson (1970), and it parallels the discussion of the headcount curves by Aaberge and Atkinson (2013). The "dual approach," based on the dual independence axiom, is analogous to the rank-dependent measurement of inequality introduced by Weymark (1981) and Yaari (1988) and to the way to summarization of the informational content of Lorenz curves by Aaberge (2001). In what follows, we draw on Aaberge and Peluso (2011) for the dual approach and Aaberge and Brandolini (2014) for the primal approach.

3.3.2.3 Complete Orderings: The Dual Approach

The dual independence axiom can be used to justify the following family of deprivation measures

$$D_\Gamma(F) = r - \sum_{k=0}^{r-1} \Gamma\left(\sum_{j=0}^{k} q_j\right) = \begin{cases} \mu + \Delta_\Gamma(F) & \text{when } \Gamma \text{ is convex} \\ \mu - \Delta_\Gamma(F) & \text{when } \Gamma \text{ is concave} \end{cases}, \tag{3.6}$$

where

$$\Delta_\Gamma(F) = \begin{cases} \sum_{k=0}^{r-1}\left[\sum_{j=0}^{k} q_j - \Gamma\left(\sum_{j=0}^{k} q_j\right)\right] & \text{when } \Gamma \text{ is convex} \\ \sum_{k=0}^{r-1}\left[\Gamma\left(\sum_{j=0}^{k} q_j\right) - \sum_{j=0}^{k} q_j\right] & \text{when } \Gamma \text{ is concave} \end{cases}, \tag{3.7}$$

and Γ, with $\Gamma(0)=0$ and $\Gamma(1)=1$, is a nonnegative, nondecreasing continuous function. Because F denotes the distribution of the deprivation count, $D_\Gamma(F)$ can be treated as a summary measure of deprivation exhibited by the distribution F. It can be seen as the social evaluation function corresponding to the social preference relation that identifies the most favorable distribution F with the one that minimizes $D_\Gamma(F)$. These social preferences are shaped by the specification of the function Γ, which can be considered as a deprivation intensity function. $D_\Gamma(F)$ can be decomposed into the mean number of deprivations, μ, and a term that captures the dispersion of deprivations across the population, Δ_Γ. By definition, Δ_Γ is always nonnegative and measures left-tail heaviness (left-spread) when Γ is concave and right-tail heaviness (right-spread) when Γ is convex. It follows that $\mu \le D_\Gamma(F) \le r$ when Γ is convex, and $0 \le D_\Gamma(F) \le \mu$ when Γ is concave. If Γ is convex, the minimum value μ of $D_\Gamma(F)$ is attained when $\Delta_\Gamma(F)=0$, that is, when each individual suffers from the same number μ of deprivations. If everybody suffers from all r deprivations, $\Delta_\Gamma(F)$ still equals 0, but $D_\Gamma(F)$ reaches its maximum value r. Conversely, $\Delta_\Gamma(F)$ is at a maximum when half of the population does not suffer from any deprivation and the remaining half suffer from all, so that $D_\Gamma(F)=r[1-\Gamma(0.5)]$. The comparison between the last two cases illustrates how the index works: a situation in which everybody suffers from r deprivations is definitely worse than one where only half of the population suffers from r deprivations. But the extent to which the two situations are valued differently depends on the convexity of Γ: the more convex it is, the more weight we give to multiple deprivations, and the closer $D_\Gamma(F)$ is to r. A similar reasoning applies, *mutatis mutandis*, for concave Γ.

Expression (3.6) shows that an exclusive concern for the mean number of deprivations implies linear (both convex and concave) social preferences: $\Gamma(t)=t$. There is indifference between a situation in which s people have one deprivation and a situation in which only one person is deprived but in s dimensions. It is the same result that we would obtain by applying the composite index approach discussed in Section 3.3.1, and it is another way to appreciate the restrictions imposed on social preferences in that approach. When there is a concern for the distribution of deprivations across the population, the critical judgement is whether this concern should prioritize the intensity or the diffusion of deprivations. In the former case, social preferences pay more attention to one person with s deprivations than to s people with one deprivation each, and the measure D_Γ should embody a convex Γ. In the latter case, social preferences take the opposite stance, and the measure D_Γ should embody a concave Γ. With Γ concave, for a given μ, D_Γ decreases as Δ_Γ increases because the distribution of deprivations across the population shifts towards people with none or fewer deprivations or, in other words, to the left tail of the distribution.

Thus, there is a correspondence between convexity and the intersection criterion on one side and between concavity and the union criterion on the other. This can be seen by taking particular specifications of the function Γ. With the union criterion, the focus is on the proportion of people who suffer from deprivation in at least one dimension $(1-q_0)$. By specifying Γ as

$$\Gamma(t)=\begin{cases} q_0 & \text{if } t=q_0 \\ 1 & \text{if } q_0<t\leq 1 \end{cases}, \tag{3.8}$$

we get $D_\Gamma(F)=1-q_0$, which means that the union measure can be considered as a limiting case of the D_Γ-family of deprivation measures in the concave case. With the intersection criterion, the focus is on the proportion of people deprived in all dimensions (q_r). The following alternative specification for Γ,

$$\Gamma(t)=\begin{cases} 0 & \text{if } 0\leq t<1-q_r \\ 1-q_r & \text{if } 1-q_r\leq t\leq 1 \end{cases}, \tag{3.9}$$

yields $D_\Gamma(F)=r-1+q_r$, which means that the intersection measure also represents a limiting case of the D_Γ-family of deprivation measures in the convex case. Although the union and intersection measures do not belong to the D_Γ-family, which is generated by continuous Γ functions, they can be approximated within this class (see Le Breton and Peluso, 2010, for general approximation results).

A Gini version of the measure of deprivation D_Γ can be obtained by taking $\Gamma(t) = 2t - t^2$ (concave) or $\Gamma(t) = t^2$ (convex), so that Δ_Γ equals the Gini mean difference. A general family of deprivation measures associated with the Lorenz family of inequality measures (Aaberge, 2000) is obtained by using the specification $\Gamma(t) = t^\tau$ for which the parameter $\tau > 0$ captures the concern for deprivation inequality, paying more attention to the lower tail when $0 < \tau < 1$ and to the upper tail when $\tau > 1$ (Aaberge and Peluso, 2011).

3.3.2.4 Complete Orderings: The Primal Approach

The independence axiom provides a justification for the following alternative family of deprivation measures,

$$d_\gamma(F) = \sum_{k=0}^{r} \gamma(k) q_k = \begin{cases} \gamma(\mu) + \delta_\gamma(F) & \text{when}\,\gamma\,\text{is}\,\text{convex} \\ \gamma(\mu) - \delta_\gamma(F) & \text{when}\,\gamma\,\text{is}\,\text{concave} \end{cases}, \tag{3.10}$$

where

$$\delta_\gamma(F) = \begin{cases} \displaystyle\sum_{k=0}^{r} (\gamma(k) - \gamma(\mu)) q_k & \text{when}\,\gamma\,\text{is}\,\text{convex} \\ \displaystyle\sum_{k=0}^{r} (\gamma(\mu) - \gamma(k)) q_k & \text{when}\,\gamma\,\text{is}\,\text{concave} \end{cases}, \tag{3.11}$$

and $\gamma(k)$, with $\gamma(0) = 0$, is a nonnegative, non-decreasing continuous function of the number of deprivations k. As with Γ in the dual case, γ can be considered as a deprivation intensity function with a curvature that determines how much we dislike increasingly severe deprivations in the convex case or growingly diffuse deprivations in the concave case. This family of deprivation measures is analogous to the family of inequality measures introduced by Kolm (1969) and Atkinson (1970). Chakravarty and D'Ambrosio (2006) provide an alternative axiomatic justification of (3.10) with a convex γ for measuring social exclusion.[22]

[22] Unlike the discussion in this section, Chakravarty and D'Ambrosio (2006) focus on the distribution of deprivation scores across people rather than on the distribution of deprivation scores itself. They also prove that second-degree downward dominance implies a convex γ and is preserved under a "favorable composite change," which is an intervention principle that is closely related to the Pigou–Dalton principle of transfers. This principle differs from the association-rearrangement principles motivated by the measurement of multidimensional poverty and discussed in Section 3.3.2.5. The index characterized by Bossert et al. (2013) is a special (linear) case of Chakravarty and D'Ambrosio's social exclusion measure. In a related paper, Bossert et al. (2007) use the counting approach to derive a further measure of social exclusion. They define axioms such that the degree of deprivation of an individual is proportional to the product of the share of people who suffer from fewer deprivations than he does and the mean difference between his deprivation score and that of all people who are better off: summation of these individual functions across individuals and then over time yields the aggregate deprivation and social exclusion indices, respectively.

As in the dual case, the primal measures $d_\gamma(F)$ can be considered as a social evaluation function for which preferences favor the count distribution F that minimizes $d_\gamma(F)$. The primal measures $d_\gamma(F)$ can be decomposed into a first term that is a transformation of the mean μ and a second term $\delta_\gamma(F)$ that measures the left- or right-tail heaviness when γ is concave or convex. By inserting $\gamma(k)=2rk-k^2$ (concave) and $\gamma(k)=k^2$ (convex) in (3.11), the term $\delta_\gamma(F)$ equals the variance. When $\gamma(k)=k$ for all k, $d_\gamma(F)=\mu$ and only the mean matters: social preferences ignore the dispersion of deprivations.[23] When the dispersion matters, as in the dual case, the judgement depends on whether social preferences give more weight to s people with one deprivation each or to one person with s deprivations, which means choosing a concave function γ in the first case and a convex function in the second. Indeed, the union criterion is a limiting case of the d_γ -family of deprivation measures for concave γ, and the intersection criterion is a limiting case for convex γ.[24] With a concave γ, the dispersion term is subtracted from the (transformed) mean and $0 \le d_\gamma(F) \le \gamma(\mu)$, whereas with a convex γ, the opposite happens, and $\gamma(\mu) \le d_\gamma(F) \le \gamma(r)$.

Unlike the dual measures, the primal measures are exactly decomposable by population subgroups, in the sense that the index computed for the overall population equals the weighted average of the measures calculated for each subgroup, with weights equal to the respective population shares of the subgroups. Note, that the dual measures may admit a different decomposition into within-group and between-group components, however, along the lines suggested by Ebert (2010).

The measure d_γ generalizes the counting measure proposed by Atkinson (2003, p. 62) for a bivariate distribution ($r=2$). Atkinson's measure A_θ can be written as

$$\begin{aligned} A_\theta &= 2^{-\theta}\left[p_{1+}+p_{+1}+2\left(2^{\theta-1}-1\right)p_{11}\right] = 2^{-\theta}(p_{1+}+p_{+1})+\left(1-2^{1-\theta}\right)p_{11} \\ &= 2^{-\theta}q_1+q_2, \end{aligned} \tag{3.12}$$

by making use of the notation of Table 3.1 and after dividing through the original formula by 2^θ. We can obtain (3.12) from (3.10) by inserting $\gamma(k)=(k/r)^\theta$ and $r=2$. The parameter θ varies from 0 to infinity and is introduced by Atkinson to capture alternative views on the importance of multiple deprivations. (Strictly speaking, both extreme values

[23] As seen, both D_Γ and d_γ can coincide with the mean μ for certain specifications of social preferences ($\Gamma(t)=t$ and $\gamma(k)=k$). From the proof of Theorem 5 in Aaberge (2001), it follows that the mean is the only measure of deprivation that satisfies both the independence and the dual independence axioms. Thus, the independence and the dual independence axioms provide, together with the conditions of transitivity, completeness, continuity, and first-degree dominance, a complete axiomatic characterization of the mean μ. In the alternative axiomatic justification for the mean offered by Bossert et al. (2013), two conditions of subgroup decomposability play a similar role as the two independence axioms.

[24] This can be seen by approximating the concave function γ with $\gamma(k)=1$ for $k=1,2,\ldots,r$ and the convex function γ with $\gamma(k)=0$ for $k=1,2,\ldots,r-1$ and $\gamma(r)=1$, which yield $d_\gamma(F)=1-q_0$ and $d_\gamma(F)=q_r$, respectively.

are inconsistent with the assumed continuity of the function γ, and they should be seen as limiting cases.) When $\theta \to 0$, the index counts all people with at least one deprivation, regardless of their number for each individual: $A_0 = p_{1+} + p_{+1} - p_{11} = q_1 + q_2$. When $\theta = 1$, people with two deprivations are counted twice and A_1 gives the simple mean of the headcount rates in the two dimensions, providing the same result as would be generated with a composite index. As θ goes to infinity, the index tends to coincide with the proportion of people deprived on both dimensions: $A_\infty \to p_{11}$. As the original Atkinson's counting deprivation index, its generalization to more than two dimensions obtained by inserting $\gamma(k) = (k/r)^\theta$ in (3.10) embodies, as limiting cases, both the union criterion (A_0) and the intersection criterion (A_∞). This index characterizes a family of deprivation measures that may be seen as the analog of the poverty measures proposed by Foster et al. (1984), referred to as the *FGT* measures.

The decomposition of the primal and dual measures of deprivation in terms of mean (or transformation of the mean) and dispersion of the deprivation count distributions parallels the mean-inequality decomposition of the social welfare functions derived from the expected and rank-dependent utility-like theories (see Atkinson, 1970; Yaari, 1988). Unlike the income inequality analysis, the structure of the decomposition of the deprivation measures depends on whether social preferences are associated with the union or the intersection criterion, however. In the former case, the deprivation measures fall and social welfare rises when the dispersion of deprivation across the population goes up, meaning that more people are affected by few or no deprivations. Even though they allow for the decomposition in terms of mean dispersion of deprivation, the primal and dual summary measures are silent about the role played by each dimension. Thus, the information provided by these summary measures should be complemented with estimates of the proportions of people who suffer from deprivation in each of the dimensions. This information reveals whether deprivation is concentrated on few or many dimensions.

Table 3.3 shows the estimates for some deprivation indices for the five European countries considered earlier. (Some indices are discussed in the next sections.) As regards dual measures, we consider the class of indices associated with the Lorenz family of inequality measures,

$$D_\tau^{\mathrm{GG}} = r - \sum_{k=0}^{r-1} \left(\sum_{j=0}^{k} q_j \right)^\tau$$

for various values of the parameters τ. For $\tau = 2$, the previous expression gives the convex version of the Gini-type measure of deprivation, and the concave version is given by:

$$D_2^{G,\,\mathrm{concave}} = 2\mu - r + \sum_{k=0}^{r-1} \left(\sum_{j=0}^{k} q_j \right)^2 = 2\mu - D_2^{\mathrm{GG}} = 2\mu - D_2^{G,\,\mathrm{convex}}.$$

Table 3.3 Indices of material deprivations in selected European countries in 2012

Index		Germany	France	Italy	United Kingdom	Norway	Germany versus France (1)	United Kingdom versus Italy (1)
Linear indices								
Mean deprivations		0.822	0.877	1.471	1.109	0.320	−6.3	−24.6
Mean headcount ratio		0.091	0.097	0.163	0.123	0.036	−6.3	−24.6
Concave indices								
D_τ^{GG}	$\tau = 0.1$	0.096	0.103	0.191	0.136	0.034	-7.3	−28.7
	$\tau = 0.5$	0.446	0.479	0.845	0.619	0.165	−6.8	−26.7
	$\tau = 0.9$	0.752	0.803	1.360	1.020	0.290	−6.4	−25.0
$D_2^{G,\text{concave}}$		0.231	0.262	0.629	0.394	0.037	−11.7	−37.4
d_θ^{GA}	$\theta \to 0$	0.400	0.420	0.604	0.510	0.166	−4.8	−15.6
	$\theta = 0.1$	0.340	0.358	0.523	0.436	0.140	−5.0	−16.6
	$\theta = 0.5$	0.184	0.195	0.303	0.241	0.074	−5.7	−20.4
	$\theta = 0.9$	0.104	0.111	0.184	0.140	0.041	−6.2	−23.8
$d_2^{V,\text{concave}}$		12.550	13.399	21.883	16.747	4.914	−6.3	−23.5
Convex indices								
D_τ^{GG}	$\tau = 1.1$	0.890	0.948	1.576	1.195	0.350	−6.2	−24.2
	$\tau = 5$	2.453	2.537	3.460	2.942	1.280	−3.3	−15.0
	$\tau = 19$	3.906	3.910	4.612	4.368	2.799	−0.1	−5.3
	$\tau = 21$	4.003	3.998	4.673	4.461	2.917	0.1	−4.5
	$\tau = 40$	4.581	4.522	5.020	5.011	3.629	1.3	−0.2
	$\tau = 42$	4.622	4.559	5.044	5.050	3.680	1.4	0.1
	$\tau = 100$	5.272	5.145	5.414	5.670	4.505	2.5	4.7
$D_2^{G,\text{convex}} = D_2^{GG}$		1.413	1.492	2.313	1.824	0.603	−5.3	−21.1
d_θ^{GA}	$\theta = 1.1$	0.080	0.086	0.146	0.109	0.031	−6.3	−25.3
	$\theta = 2$	0.028	0.029	0.057	0.040	0.010	−5.9	−30.0
	$\theta = 3$	0.011	0.011	0.023	0.016	0.004	−3.6	−31.6
	$\theta = 4$	0.005	0.005	0.010	0.007	0.002	0.4	−30.1
	$\theta = 8$	0.001	0.001	0.001	0.001	0.000	20.6	−13.5
	$\theta = 9$	0.0003	0.0002	0.0005	0.0005	0.0001	42.8	2.3
	$\theta = 20$	7.6×10^{-06}	1.3×10^{-06}	7.8×10^{-06}	9.4×10^{-06}	6.6×10^{-06}	479.9	20.9
$d_2^{V,\text{convex}} = r^2 d_2^{GA}$		2.246	2.387	4.595	3.215	0.846	−5.9	−30.0
Other indices								
Eurostat SMD (2)		0.049	0.052	0.145	0.079	0.017	−5.8	−45.5

Note: (1) Percentage relative deviation of the figure for the first country from the figure for the second country. (2) Figures are computed from Table 3.2 and may differ from published statistics because of rounding.
Source: authors' elaboration on data from Eurostat (2014).

In regard to primal measures, we consider the generalized Atkinson-type class of indices

$$d_\theta^{\mathrm{GA}} = r^{-\theta}\sum_{k=1}^{r} k^\theta q_k$$

for various values of the parameters θ. For $\theta=1$, the previous expression gives the mean headcount ratio, which equals the ratio μ/r. For $\theta=2$, it coincides with the convex version of the variance-type measure of deprivation $d_2^{V,\mathrm{convex}}$ multiplied by r^{-2}, and the concave version ($\gamma(k)=2rk-k^2$) is given by:

$$d_2^{V,\mathrm{concave}} = 2r\mu - \sum_{k=1}^{r} k^2 q_k = 2r\mu - r^2 d_2^{\mathrm{GA}} = 2r\mu - d_2^{V,\mathrm{convex}}.$$

Norway shows the lowest mean number of deprivations followed by Germany and France, which are rather close each other, the United Kingdom, and finally Italy. The mean headcount ratio ranges from 3.6% in Norway to 16.3% in Italy. With a concave index, we always find that deprivation is lower in Germany than in France and lower in the United Kingdom than in Italy, which is not surprising in light of the results on second-degree upward dominance reported in Section 3.3.2.1. On the other hand, the lack of second-degree downward dominance in these same comparisons is noticeable in the fact that the rankings reverse as the functions become more convex. For instance, the generalized Atkinson-type deprivation index turns out to be lower in France than in Germany for values of θ higher than 4. The French overall deprivation is below the German level whenever we favor the intersection criterion and weight somebody suffering from $2h$ deprivations at least 16 ($=2^4$) times somebody suffering from h deprivations (as the index d_4^{GA} assigns each person with h deprivations a weight equal to h^4). Because the United Kingdom fares much better than Italy, except in the occurrence of very severe deprivation (6 or more items), the ranking between the two countries only changes for high values of θ or τ, which correspond to an extreme aversion to the worst conditions of deprivation. Finally, note that the generalized Atkinson-type deprivation index approaches the proportion of people experiencing at least one deprivation (union criterion) as θ tends to 0, and the proportion of people suffering from the maximum number of deprivations (intersection criterion) as θ goes to infinity; as nobody lacks all nine items, in the latter case, the index converges to zero in all countries.

3.3.2.5 Association Rearrangements

In many respects, the discussion so far has proceeded as in the case of a single variable, whereas the key feature of the multivariate case is the pattern of association across dimensions. It is then natural to ask how social welfare responds to a change in the

Table 3.4 Illustration of a marginal-free positive association-increasing rearrangement

	$X_2=0$	$X_2=1$	
$X_1=0$	0.35	0.20	0.55
$X_1=1$	0.20	0.25	0.45
	0.55	0.45	1

	$X_2=0$	$X_2=1$	
$X_1=0$	0.36	0.19	0.55
$X_1=1$	0.19	0.26	0.45
	0.55	0.45	1

Source: authors' elaboration.

Table 3.5 Illustration of a marginal-free negative association-increasing rearrangement

	$X_2=0$	$X_2=1$	
$X_1=0$	0.20	0.25	0.45
$X_1=1$	0.35	0.20	0.55
	0.55	0.45	1

	$X_2=0$	$X_2=1$	
$X_1=0$	0.19	0.26	0.45
$X_1=1$	0.36	0.19	0.55
	0.55	0.45	1

Source: authors' elaboration.

distribution of deprivations across the population, although the total number of deprivations remains the same. The standard approach is to consider how social welfare varies after a "marginal-free change" in the association between two variables, which is a change that does not affect the marginal distributions. As in the statistical literature on the measurement of association in multidimensional contingency tables (formed by two or several binary variables), we distinguish association rearrangements for distributions characterized by either positive or negative associations. Illustrations of marginal-free association rearrangements are provided by Tables 3.4 and 3.5. Each panel of Table 3.4 is obtained from the opposite panel by a marginal-free positive association-increasing (decreasing) rearrangement, whereas each panel of Table 3.5 can be obtained from the opposite panel by a negative association-increasing (decreasing) rearrangement.

Marginal-free rearrangements have been widely used as a basis for evaluating multidimensional measures of poverty and inequality.[25] Bourguignon and Chakravarty (1999, 2003, 2009) and Atkinson (2003) use the principle of marginal-free correlation-increasing shifts as a basis for making a normative judgement of poverty measures derived

[25] For definitions of association-increasing rearrangements based on the correlation coefficient, we refer to Epstein and Tanny (1980), Atkinson and Bourguignon (1982), Boland and Proschan (1988), Dardanoni (1995), Tsui (1995, 1999, 2002), Bourguignon and Chakravarty (2003), and Duclos et al. (2006a). See also Tchen (1980) who deals with positive association (or concordance) between bivariate probability measures.

from continuous variables (attributes) rather than from deprivation scores. They distinguish whether the poverty measure increases or decreases because of a correlation-increasing shift, and they consider the associated attributes to be substitutes (one attribute can compensate for the lack of the other) in the former case and to be complements in the latter.

Considering marginal-free changes is a neat way to highlight the fact that the multidimensional analysis of poverty and inequality implies making assumptions regarding the degree to which the different attributes can be substituted one for the other. In the real world, the condition of marginal-free changes may be too restrictive because policies may reduce deprivation in one dimension at the cost of increasing deprivation in another. We hence adopt a more general approach, and we require that only the mean number of deprivations but not the marginal distributions be kept fixed. (The latter implies the former, but not vice versa.) It follows that we need a measure of association that is invariant with regard to changes in the marginal distributions, unlike the correlation coefficient. This is the case of the cross-product κ introduced by Yule (1900). In the 2×2 distribution of Table 3.1, Yule's measure is defined by

$$\kappa = \frac{p_{00} p_{11}}{p_{01} p_{10}}, \tag{3.13}$$

which is invariant to the transformation $p_{ij} \to a_i b_j p_{ij}$. This association measure, together with the marginal distributions (p_{0+}, p_{1+}) and (p_{+0}, p_{+1}), provides complete information on the distribution and does not change if the marginal distributions change.[26] Note that $\kappa \in [0, \infty\rangle$, $\kappa = 1$ if X_1 and X_2 are independent, $\kappa = 0$ if there is perfect negative association ($p_{00} = 0$ and/or $p_{11} = 0$), and $\kappa \to \infty$ if there is perfect positive association ($p_{01} = 0$ and/or $p_{10} = 0$).

Following Aaberge and Peluso (2011) and Aaberge and Brandolini (2014), we relax the marginal-free condition by introducing an association-increasing/decreasing rearrangement principle that relies on the condition of a fixed overall mean number of deprivations rather than on the condition of fixed proportions of people suffering from each deprivation. As illustrated by Tables 3.4 and 3.5, marginal-free arrangements are special cases of this alternative rearrangement principle.[27]

[26] Yule's measure of association is related to the copula-based measures of association for continuous variables introduced by Spearman and Kendall; see Nelsen (1998). Decancq (2014) introduces a copula-based generalization of the rearrangement principles for continuous variables and provides an analysis of their links with stochastic dominance. If X_1 and X_2 represent the two social class categories to which an individual can belong at times 1 and 2, the Yule's measure of association also coincides with the odds ratio used in mobility studies. See, for instance, Erikson and Goldthorpe (1993, p. 55) and Chapter 10.

[27] Note that the multinomial distribution defined by the parameters p_{00}, p_{10}, p_{01} and $p_{11} (= 1 - p_{00} - p_{10} - p_{01})$ can alternatively be described by the marginal distributions $(p_{0+}, p_{1+} = 1 - p_{0+})$ and $(p_{+0}, p_{+1} = 1 - p_{+0})$, and the cross-product κ.

Definition 3.3

Consider a 2×2 *table with parameters* $(p_{00}, p_{01}, p_{10}, p_{11})$ *where* $\sum_i \sum_j p_{ij} = 1$. *The change* $(p_{00} + \varepsilon, p_{01}, p_{10} - 2\varepsilon, p_{11} + \varepsilon)$ *is said to provide a mean-preserving positive association-increasing (decreasing) rearrangement if* $\varepsilon > 0$ $(\varepsilon < 0)$ *and* $\kappa > 1$, *and a mean-preserving negative association-increasing (decreasing) rearrangement if* $\varepsilon < 0$ $(\varepsilon > 0)$ *and* $\kappa < 1$.

It follows from Definition 3.3 that a mean-preserving rearrangement reduces the number of people deprived according to indicator X_1 at the cost of increasing the number of people deprived according to indicator X_2 when $\varepsilon > 0$ and vice versa when $\varepsilon < 0$. This is illustrated in Table 3.6, which shows two distributions where the association is negative $(\kappa < 1)$ and the mean is equal to 1. Each panel can be obtained from the opposite panel by a mean-preserving negative association-decreasing (increasing) rearrangement where $\varepsilon = 0.01$.

Aaberge and Peluso (2011) show how to extend Definition 3.3 to r dimensions. As is the standard, subscript notation becomes cumbersome for more than two dimensions, so they simplify the notation to p_{ijm}, where i and j represent two arbitrary chosen deprivation dimensions and m represents the remaining $r-2$ dimensions. The Yule's measure κ_{ijm} is defined by

$$\kappa_{ijm} = \frac{p_{iim} p_{jjm}}{p_{ijm} p_{jim}} \tag{3.14}$$

where m is an $(r-2)$-dimensional vector of any combination of zeroes and ones. In this case, the association is defined by $r(r-1)/2$ cross-products. Aaberge and Peluso (2011) introduce the following generalization of Definition 3.3:

Definition 3.4A

Consider a $2 \times 2 \times \ldots \times 2$ *table formed by s dichotomous variables with parameters* $(p_{iim}, p_{ijm}, p_{jim}, p_{jjm})$ *where* $\sum_i \sum_j \sum_m p_{ijm} = 1$ *and* $\kappa_{ijm} > 1$. *The following change* $(p_{iim} + \varepsilon, p_{ijm}, p_{jim} - 2\varepsilon, p_{jjm} + \varepsilon)$ *is said to provide a mean-preserving positive association-increasing (decreasing) rearrangement if* $\varepsilon > 0$ $(\varepsilon < 0)$.

Definition 3.4B

Consider a $2 \times 2 \times \ldots \times 2$ *table formed by s dichotomous variables with parameters* $(p_{iim}, p_{ijm}, p_{jim}, p_{jjm})$ *where* $\sum_i \sum_j \sum_m p_{ijm} = 1$ *and* $\kappa_{ijm} < 1$. *The following change* $(p_{iim} + \varepsilon, p_{ijm}, p_{jim} - 2\varepsilon,$

Table 3.6 Illustration of a mean-preserving negative association-decreasing rearrangement

	$X_2 = 0$	$X_2 = 1$	
$X_1 = 0$	0.20	0.30	0.50
$X_1 = 1$	0.30	0.20	0.50
	0.50	0.50	1

	$X_2 = 0$	$X_2 = 1$	
$X_1 = 0$	0.21	0.30	0.51
$X_1 = 1$	0.28	0.21	0.49
	0.49	0.51	1

Source: authors' elaboration.

$p_{jjm}+\varepsilon$) is said to provide a mean-preserving negative association-increasing (decreasing) rearrangement if $\varepsilon<0$ ($\varepsilon>0$).

Theorem 3.1A demonstrates that social preferences favoring second-degree downward dominance imply that overall deprivation rises after a mean-preserving positive association-increasing rearrangement, as well as a mean-preserving negative association-decreasing rearrangement, irrespective of whether preferences are consistent with the primal or the dual approach. By contrast, Theorem 3.1B proves that preferences favoring upward second-degree dominance consider such rearrangement as a reduction in the overall deprivation. Moreover, it follows directly from the decompositions (3.6) and (3.10) that the principles of mean-preserving association-increasing/decreasing rearrangement are equivalent to the mean-preserving spread/contraction defined by:

Definition 3.5
Let F_1 and F_2 be members of the family $\boldsymbol{F}$ of count distributions based on r deprivations and assume that they have equal means. Then F_2 is said to differ from F_1 by a mean-preserving spread (contraction) if $\Delta_\Gamma(F_2)>\Delta_\Gamma(F_1)$ for all convex Γ or $\delta_\gamma(F_2)>\delta_\gamma(F_1)$ for all convex γ ($\Delta_\Gamma(F_2)<\Delta_\Gamma(F_1)$ for all concave Γ or $\delta_\gamma(F_2)<\delta_\gamma(F_1)$ for all concave γ).

Note that Definition 3.5 is equivalent to a sequence of the mean-preserving spread introduced by Rothschild and Stiglitz (1970).

Let Ω_1 and Ω_2 be subsets of the Γ-family defined by

$$\Omega_1=\{\Gamma:\Gamma'(t)>0,\Gamma''(t)>0\ \text{ for all } t\in\langle 0,1],\ \text{ and } \Gamma'(0)=0\}$$

and

$$\Omega_2=\{\Gamma:\Gamma'(t)>0,\Gamma''(t)<0\ \text{ for } t\in\langle 0,1\rangle,\ \text{ and } \Gamma'(1)=0\},$$

and let ω_1 and ω_2 be subsets of the γ-family defined by

$$\omega_1=\{\gamma:\gamma'(k)>0,\gamma''(k)>0\ \text{ for all } k>0,\ \text{ and } \gamma'(0)=0\}$$

and

$$\omega_2=\{\gamma:\gamma'(k)>0,\gamma''(k)<0\ \text{ for } k>0,\ \text{ and } \gamma'(r)=0\}.$$

All members of the sets Ω_1 and ω_1 are increasing convex functions, and all members of Ω_2 and ω_2 are increasing concave functions.

Theorem 3.1A
Let F_1 and F_2 be members of the family $\boldsymbol{F}$ of count distributions based on r deprivations and assume that they have equal means. Then the following statements are equivalent:

(i) *F_1 second-degree downward dominates F_2;*
(ii) *$D_\Gamma(F_1) < D_\Gamma(F_2)$ for all $\Gamma \in \Omega_1$;*
(iii) *$d_\gamma(F_1) < d_\gamma(F_2)$ for all $\gamma \in \omega_1$;*
(iv) *F_2 can be obtained from F_1 by a sequence of mean-preserving positive association-increasing rearrangements when $\kappa > 1$ for both F_1 and F_2, a sequence of mean-preserving negative association-decreasing rearrangements when $\kappa < 1$ for both F_1 and F_2, and a combination of mean-preserving positive association-increasing and negative association-decreasing rearrangements when $\kappa > 1$ for eitherF_1 or F_2;*
(v) *F_2 can be obtained from F_1 by a mean-preserving spread.*

Theorem 3.1B

Let F_1 and F_2 be members of the family $\boldsymbol{F}$ of count distributions based on r deprivations and assume that they have equal means. Then the following statements are equivalent:
(i) *F_1 second-degree upward dominates F_2;*
(ii) *$D_\Gamma(F_1) < D_\Gamma(F_2)$ for all $\Gamma \in \Omega_2$;*
(iii) *$d_\gamma(F_1) < d_\gamma(F_2)$ for all $\gamma \in \omega_2$;*
(iv) *F_2 can be obtained from F_1 by a sequence of mean-preserving positive association-decreasing rearrangements when $\kappa > 1$ for both F_1 and F_2, a sequence of mean-preserving negative association-increasing rearrangements when $\kappa < 1$ for both F_1 and F_2, and a combination of mean-preserving positive association-decreasing and negative association-increasing rearrangements when $\kappa > 1$ for either F_1 or F_2;*
(v) *F_2 can be obtained from F_1 by a mean-preserving contraction.*

See Aaberge and Peluso (2011) for a proof of the equivalence between (i), (ii), and (iv) of Theorems 3.1A and 3.1B, and Aaberge and Brandolini (2014) for a proof of the equivalence between (i) and (iii). The equivalence between (v) and (ii) and (iii) follows directly from the second terms of Equations (3.6) and (3.10).

Following the distinction made by Bourguignon and Chakravarty (2003, 2009) and Atkinson (2003), the results of Theorem 3.1A (3.1B) justify the use of D_Γ and d_γ for convex Γ and convex γ (concave Γ and concave γ) when the attributes associated with the deprivation indicators can be considered as substitutes (complements). Theorems 3.1A and 3.1B show that D_Γ and d_γ satisfy the mean-preserving association-rearrangement principles, when a distinction has been made between whether an association rearrangement comes from a distribution characterized by a positive or negative association. Consider the specific subfamily of two-dimensional deprivation measures discussed by Atkinson (2003) and defined by (3.12), and assume that there is positive association between the two deprivations ($\kappa > 1$). The d_γ-function associated with the family A_θ is concave for $\theta < 1$ and convex for $\theta > 1$, and it approaches the union condition when $\theta \to 0$ and the intersection condition when $\theta \to \infty$. Theorem 3.1B states that a sequence of mean-preserving positive association-decreasing rearrangements raises the overall deprivation A_θ if $\theta < 1$. Is it reasonable to suppose that the overall deprivation rises as we observe a reduction in the positive association between deprivations in the two attributes? After all, the share of people suffering from deprivation for both attributes falls, while the total number of deprivations does not vary. The answer is positive if we regard the two attributes as complements, which means that we rule out any trade-off between them, and we dislike the fact that more people are deprived more than the fact that fewer people are hit more.

Until now, we have not considered the cases of unequal weighting of the dimensions. However, all results summarized by Theorems 3.1A and 3.1B remain valid for the distribution of weighted deprivation counts. For the dual approach, Aaberge and Peluso (2011) account for different weights by considering the weighted deprivation counts $X = \sum_{i=1}^{r} w_i X_i$ and the associated distribution F, for which $w_1 \leq w_2 \leq \cdots \leq w_r$. For the primal approach, we could apply the procedure suggested by Alkire and Foster (2011a,b) to replace the deprivation count for each person with the sum of the associated weights.

3.3.2.6 Counting Deprivations versus Measuring Poverty

So far, we have been concerned with the distribution of deprivation counts, irrespective of how many people are regarded as poor when deprivation and poverty are considered as distinct concepts. In terms of the classical distinction made by Sen (1976), we have focused only on the aggregation of the characteristics of deprivation into an overall measure of deprivation, ignoring the first step concerning the identification of the poor. The contrast between the union and the intersection criteria emphasized in the previous sections suggests, however, that there is some leeway in defining who is poor. For instance, Brandolini and D'Alessio (1998), Bourguignon and Chakravarty (1999, 2003), Tsui (2002), and Bossert et al. (2013) adopt the more extensive union criterion and define people as (multidimensional) poor if they suffer from at least one deprivation. In this case, deprivation and poverty come to coincide. On the other hand, the European Union regards as multiply materially deprived all persons who cannot afford at least four out of nine amenities, moving midway between the union and the (strict) intersection views. Alkire and Foster (2011a,b) formalize what they label the "dual cut-off" identification system, in which the dimension-specific thresholds are integrated with a further threshold that identifies the minimum number of deprivations required for an individual to be classified as poor. If a person is poor when he or she is deprived in at least c, $1 \leq c \leq r$, dimensions, the headcount ratio is uniquely determined by the count distribution F and is defined by

$$\widetilde{H}(c) = 1 - F(c-1) = \sum_{k=c}^{r} q_k. \tag{3.15}$$

In the case of the European indicator of severe material deprivation, c equals 4. As the choice of a specific cut-off c is arbitrary, it is useful to check the sensitivity of the ranking of distributions to c by treating $\widetilde{H}(c)$ as a function of c, henceforth labeled the headcount curve. As is evident from (3.15), the condition of first-degree dominance of headcount curves is equivalent to first-degree dominance of the associated count distributions. If $c>1$, first-degree dominance for headcount curves is a less demanding condition than that for the overall count distribution, because it ignores what happens to those who suffer from deprivation in fewer than c dimensions. Moreover, the second-degree dominance results of Theorems 3.1A and 3.1B are also valid for the headcount curve, which means that $\widetilde{H}(c)$ satisfies the principle of association-increasing/decreasing rearrangements when this principle is restricted to be applied among the poor.

To complement the information provided by the headcount ratio when only ordinal data are available, we may employ the measures defined by (3.6) and (3.10) as overall measures of poverty for the conditional count distribution $\widetilde{F}$ defined by

$$\widetilde{F}(k;c) = \Pr(X \leq k | X \geq c) = \frac{F(k) - F(c-1)}{1 - F(c-1)} = \frac{\sum_{j=c}^{k} q_j}{\sum_{j=c}^{r} q_j}, \quad k = c, c+1, \ldots, r, \tag{3.16}$$

with mean given by

$$\widetilde{\mu}(c) = \frac{\sum_{j=c}^{r} j q_j}{\sum_{j=c}^{r} q_j}. \tag{3.17}$$

Expressions (3.6) and (3.10) show that the overall measures of poverty for $\widetilde{F}$ admit a decomposition into the mean (or a function of the mean) and a measure of dispersion. An analog to the *FGT* family of poverty measures is obtained by inserting $\gamma(k) = k^{\theta}$ in expression (3.10).

As an alternative, Alkire and Foster (2011a) propose combining the headcount ratio $\widetilde{H}(c)$ and the conditional mean $\widetilde{\mu}(c)$, and they introduce the adjusted headcount ratio defined by

$$\widetilde{M}_1(c) = \frac{\widetilde{H}(c)\widetilde{\mu}(c)}{r} = \frac{1}{r}\sum_{j=c}^{r} j q_j, \tag{3.18}$$

which is the ratio of the total number of deprivations experienced by the poor to the maximum number of deprivations that could be experienced by the entire population. For $c=1$, the index $\widetilde{M}_1(c)$ coincides with the Atkinson-type primal measure of deprivation d_1^{GA}. Expression (3.18) can account for unequal weights for the various deprivations by simply replacing the deprivation count for each person by the sum of the associated weights. Alkire and Foster (2011a, p. 482) underline that both the identification of the poor and the adjusted headcount ratio are invariant to monotonic transformations applied to the deprivation variables and the respective thresholds. Moreover, the index $\widetilde{M}_1(c)$ increases if a poor person becomes deprived in an additional dimension (dimensional monotonicity), and it is decomposable by population subgroups. Therefore, the index can be broken down by indicator because it is the (weighted) average of the deprivations headcount ratios for each dimension computed considering only the poor at the numerator (so-called "censored headcount ratios"). On the other hand, this index is indifferent to changes in the way deprivations are distributed across the poor.[28]

A general family of adjusted poverty measures that take into account not only the average deprivation experienced by the poor, $\widetilde{\mu}(c)$, but also the distribution of deprivations across the poor can be derived from the d-measure defined by (3.10)

$$\widetilde{M}_\gamma(c) = \frac{\widetilde{H}(c)\widetilde{d}_\gamma(c)}{r}, \tag{3.19}$$

where $\widetilde{d}_\gamma(c)$ denotes the d-index for $\widetilde{F}$. Such a measure may weight poor persons differently according to the number of deprivations from which they suffer. Inserting $\gamma(k) = k^\theta$ in $\widetilde{d}_\gamma(c)$ in (3.19) yields the general family of adjusted FGT measures for count data

$$\widetilde{M}_\theta(c) = \frac{1}{r}\sum_{j=c}^{r} j^\theta q_j, \quad \theta > 0, \tag{3.20}$$

which encompasses (3.18) for $\theta = 1$. When $\theta \to 0$, the adjusted FGT measure reaches its minimum value $\widetilde{H}(c)/r$, which ignores altogether any cumulative effect of multiple deprivations. As θ rises, greater weight is placed on those who suffer from deprivation in several dimensions.

Figure 3.3 compares how poverty headcount ratios change as we vary the poverty cut-off using the deprivation indicators in the five European countries considered earlier. The proportion of poor people, shown in the top-left panel, falls by three-fourths in Italy and around nine-tenths in the other countries as the poverty cut-off is raised from one deprivation (union criterion) to four deprivations (the European criterion). Censoring at four deprivations implies excluding from measured poverty a substantial fraction of the population suffering from one, two, or three deprivations: 15% in Norway and 46% in Italy, accounting for 76% and 57% of all deprivations, respectively. However, the ranking

[28] For comments and critiques of the class of multidimensional indices proposed by Alkire and Foster (2011a,b), see, among others, Birdsall (2011), Rippin (2010), Ferreira (2011), Ravallion (2011a, 2012a), Silber (2011), Thorbecke (2011), Ferreira and Lugo (2013), Duclos and Tiberti (forthcoming), and the reply by Alkire et al. (2011).

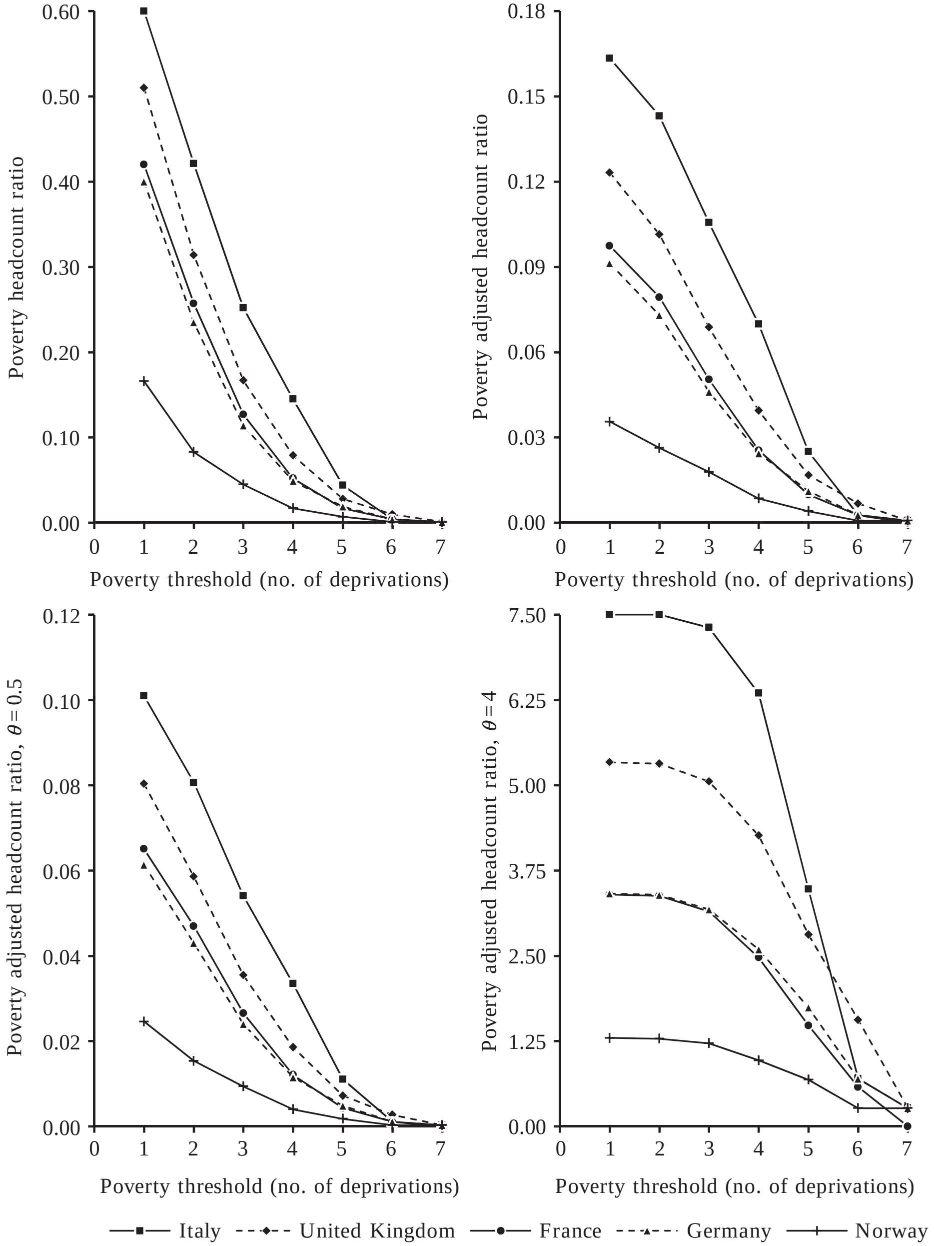


Figure 3.3 Poverty headcount and adjusted headcount ratios for different poverty cut-offs in selected European countries in 2012. *Source: authors' elaboration on data from Eurostat (2014).*

of countries does not change. It changes when the cut-off is set at five deprivations, however, when Germany and France reverse their order, and again at six deprivations, when the United Kingdom becomes the country with the highest share of poor people. In the top-right panel, the ranking is the same for the adjusted headcount ratio $\widetilde{M}_1(c)$, except for a better position granted to France by its lower average intensity of deprivation $(\widetilde{\mu}(c)/r)$ when the cut-off is set at six deprivations. The bottom panels show results for the adjusted *FGT* measure $\widetilde{M}_\theta(c)$: lowering the weights of multiple deprivations ($\theta = 0.5$; left panel) does not modify the sorting produced by the adjusted headcount ratio, whereas significantly raising them ($\theta = 4$; right panel) steadily switches the positions of Germany and France, as seen in Section 3.3.2.4. This comparison reveals that varying the poverty cut-off has a considerable impact on measured poverty, whereas adjusting the headcount ratio for the deprivations experienced by the poor seems to have minor effects, unless their distribution is taken into account.

The adjusted headcount ratio $\widetilde{M}_1(c)$ proposed by Alkire and Foster (2011a,b) provides the theoretical basis for the Multidimensional Poverty Index (*MPI*) developed by Alkire and Santos (2010, 2013, 2014).[29] The *MPI* has replaced the *HPI* in the reports of the United Nations Development Program since 2010 in order to capture "how many people experience overlapping deprivations and how many deprivations they face on average" (UNDP, 2010, p. 95). The *MPI* considers 10 dichotomous indicators for three dimensions: health, education, and living standards. Dimensions and indicators within each dimension are equally weighted, and the cut-off c for the number of (weighted) deprivations is set at three out of a maximum of ten. Applied research estimating Alkire and Foster's class of indices and the *MPI* is rapidly growing.[30]

3.3.3 Poverty Measurement Based on Continuous Variables

The counting approach focuses on the distribution of deprivation scores that summarize binary variables, defined as having or not having goods or performing or not performing activities that are seen as social necessities. When we have cardinal (continuous or categorical) variables, we can use measures of multidimensional poverty that fully exploit

[29] Alkire and Foster's method is utilized by Peichl and Pestel (2013a,b) to derive an adjusted headcount ratio for multidimensional richness. This index accounts for the number of individuals who are affluent in a minimum number of dimensions, as well as for their average achievements in these dimensions.

[30] See for instance, Roelen et al. (2010) for Vietnam, Khan et al. (2011) for Pakistan, Batana (2013) for Sub-Saharan African countries, Battiston et al. (2013) for Latin American countries, Roche (2013) for Bangladesh, Santos (2013) or Buthan, Trani, and Cannings (2013) for Western Darfur, Trani et al. (2013) for Afghanistan, Yu (2013) for China, and Cavapozzi et al. (2013) and Whelan et al. (2014) for European countries. See also Mohanty (2011) for a related study on deprivation scores in India. Bennett and Mitra (2013) develop multiple statistical tests for Alkire and Foster's family of poverty measures.

the informational richness of the available data.[31] As in the counting approach, we may aggregate attributes first across dimensions and then across individuals. This procedure corresponds to representing each individual's vector of attributes with an interpersonally comparable utility-like function and then evaluating the distribution of individual well-being using the same tools as in a univariate space. Consumer theory (Slesnick, 1993) or information theory (Maasoumi and Lugo, 2008) can provide the analytical framework to derive the utility-like function. This function is then used to aggregate the attribute-specific cut-offs to define an aggregate poverty threshold.[32]

Alternatively, we may employ an axiomatic simultaneous aggregation approach for measuring multidimensional poverty. Chakravarty et al. (1998), Bourguignon and Chakravarty (1999, 2003), and Tsui (2002) consider persons to be poor if they suffer from at least one deprivation (the union approach), whereas Alkire and Foster (2011a) take all those who are deprived in at least c dimensions, with c between 1 and r. All these papers then aggregate the individual shortfalls relative to dimension-specific cut-offs into a multidimensional poverty measure. The actual functional forms of the poverty indices are determined by the combination of chosen axioms, many of which parallel those considered in the univariate analysis (e.g., Zheng, 1997). In the next section, we selectively review these indices and illustrate some of their properties. We refer to Chakravarty et al. (1998), Bourguignon and Chakravarty (1999, 2003), Tsui (2002), and Chakravarty and Silber (2008) for proofs and further discussion of the axioms.

3.3.3.1 Measures of Multidimensional Poverty

Let $\mathbf{y}_i, i = 1, 2, \ldots, n$ denote the vector of the attributes of individual i, where $y_{ij} \geq 0$ is the achievement in dimension j, and let $\mathbf{z}$ be a vector of attribute-specific poverty thresholds. Bourguignon and Chakravarty (1999, 2003) introduce the following multidimensional analog of the FGT family of poverty indices

$$P_\theta(y; z) = \frac{1}{nr} \sum_{i=1}^{n} \sum_{j=1}^{r} a_{ij} \left(1 - \frac{y_{ij}}{z_j}\right)^{\theta_j}, \quad \theta_j > 1, \tag{3.21}$$

[31] Bosmans et al. (2013b) introduce an approach that deals with the joint aggregation of cardinal and ordinal variables. Yalonetzky (2013) derives stochastic dominance conditions for ordinal variables.

[32] Merz and Rathjen (2014a,b) apply Maasoumi's utility-like approach, estimating a CES function, to study poverty in the bidimensional space of equivalent income and leisure time. Maasoumi and Lugo (2008) show that the information-theoretic approach can embody attribute-specific thresholds if the utility-like function is replaced by a relative deprivation function the argument for which is the relative shortfall of the attribute relative to its threshold. In this case, the aggregation across individuals has to be restricted to those who are deprived in at least one dimension.

where a_{ij} is equal to the weight w_j assigned to attribute j if $y_{ij} < z_j$ and to 0 otherwise; both a_{ij} and θ_j determine the weight assigned to attribute j in the poverty index.[33]

In addition to displaying *monotonicity*, *continuity*, and *scale invariance*, the members of the family $P_\theta(y;z)$ satisfy three axioms worthy of some comments. *Subgroup decomposability* requires that overall poverty can be expressed as a weighted sum of the poverty ratios of the subgroups and implies that the poverty index is separable across individuals. Thanks to this property, the poverty index identifies an individual poverty function. The *one-dimensional Pigou–Dalton transfer principle* demands that poverty does not increase in the case of a progressive transfer of one unit of attribute j from one poor person to a poorer person. This axiom determines the additivity of (3.21) across attributes. This *factor decomposability* may be a useful feature because it allows identifying the contribution of each attribute to the overall poverty level.

The *focus* axiom highlights the greater complexity of multidimensional analysis. In the univariate income case, it simply entails that the poverty index is independent of the distribution of income among the nonpoor. In the multivariate case, it may require the poverty measure to be invariant with respect to increases in y_{ij} if $y_{ij} > z_j$ for all i, poor and nonpoor alike (*strong focus*), or only with respect to changes in the distribution of attributes among the nonpoor (*weak focus*). The stronger version of the axiom implies that a better achievement in a dimension in which an individual is not deprived cannot compensate for a below-threshold achievement in another dimension. The possibility of trading off one attribute for the other is ruled out. This is the case of (3.21), as a_{ij} equals 0 if $y_{ij} > z_j$.[34]

The family of poverty measures $P_\theta(y;z)$ is a particular specification of the more general class characterized by Chakravarty et al. (1998) and Bourguignon and Chakravarty (1999, 2003) in which the power function is replaced by a continuous nonincreasing convex function. Assuming that all attributes are positive and specifying such a function to be the negative of the logarithm, Chakravarty and Silber (2008) and Chakravarty et al. (2008) obtain the multidimensional version of the poverty index proposed by Watts (1968) and formalized by Zheng (1993):

$$P_W(y;z) = \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} a_{ij} \ln\left(\frac{z_j}{y_{ij}}\right), \tag{3.22}$$

As indicated by Bourguignon and Chakravarty (2003) and Pattanaik et al. (2011), the measures defined by (3.21) are not sensitive to association-rearrangement interventions.

[33] See Lasso de la Vega and Urrutia (2011) for an axiomatic characterization of a generalized version of $P_\theta(y;z)$.

[34] Permanyer (2014) shows how to modify most of the multidimensional poverty indices commonly considered so that they satisfy the *weak* rather than the *strong focus* axiom. Esposito and Chiappero Martinetti (2010) examine poverty indices that embody a hierarchical ordering of well-being dimensions.

This is due to the fact that $P_\theta(y;z)$ is uniquely determined by the marginal distributions of the r attributes and the associated attribute-specific poverty thresholds. By imposing an additional condition, called the *poverty focus* axiom, Alkire and Foster (2011a) provide a justification for the intermediate poverty measure analog to $P_\theta(y;z)$,

$$P_\theta(y;z,c)=\frac{1}{nr}\sum_{i=1}^{n}a_i(c)\sum_{j=1}^{r}a_{ij}\left(1-\frac{y_{ij}}{z_j}\right)^\theta,\quad \theta\geq 0, \tag{3.23}$$

where $a_i(c)=1$ if $\{j: y_{ij}<z_j\}\geq c$ and 0 otherwise. In other words, the role of $a_i(c)$ is to select only the poor individuals: these are now all people who suffer from deprivation in at least c dimensions, which, for $c>1$, is a subgroup of those deprived in any dimension considered with the union criterion. Thus, $P_\theta(y;z,1)=P_\theta(y;z)$. When one or several attributes are dichotomous variables, then $P_\theta(y;z,c)$ is only valid for $\theta=0$. In such a case, we can only use

$$P_0(y;z,c)=\frac{1}{nr}\sum_{i=1}^{n}a_i(c)\sum_{j=1}^{r}a_{ij}, \tag{3.24}$$

which is equal to the average number of deprivations (normalized by the maximum number r) experienced by the poor, or the normalized average among those individuals who suffer from at least c deprivations. Note that $P_0(y;z,c)$ is equal to $\widetilde{M}_1(c)$, defined by (3.18), when $w_j=1$ for all j. For $c=1$, $P_0(y;z,1)$ becomes equal to the average number of deprivations (relative to r) for those who suffer from at least one deprivation. As demonstrated by Alkire and Foster (2011a), $P_\theta(y;z,c)$ for $\theta\geq 1$ satisfies a multidimensional transfer principle based on a bistochastic transformation when it is only applied among the poor. Moreover, the measures defined by (3.24) satisfy the association-rearrangement principle discussed by Alkire and Foster (2011a), even though these measures are decomposable by subgroups. However, this is only true when $c>1$, and it is due to the multidimensional information captured by the counting term $a_j(c)$ of (3.23).[35] By contrast, to account for the correlation between attributes when $c=1$, Bourguignon and Chakravarty (1999, 2003) introduce a family of nonadditive poverty measures, but they limit their discussion to the two-dimensional case.[36] This subfamily of $P_\theta(y;z)$ is defined by

[35] As before, the introduction of the threshold can be criticized because it implies ignoring the condition of those who suffer from deprivation in less than c dimensions as well as because of the arbitrariness of the choice of c.

[36] For alternative families of multidimensional poverty measures and their characterizations, we refer to Kolm (1977), Chakravarty et al. (1998), Tsui (2002), Deutsch and Silber (2005), Duclos et al. (2006a, 2007, 2008), Chakravarty and Silber (2008), and Lasso de la Vega and Urrutia (2011). Diez et al. (2008) and Chakravarty and D'Ambrosio (2013) derive subgroup-decomposable multidimensional poverty indices that are unit-consistent; that is, they provide poverty rankings that are unaffected by a change in the measurement units of dimensions.

$$P^*_{\alpha,\beta}(y;z) = \frac{1}{nr}\sum_{i=1}^{n}\left(\sum_{j=1}^{2} a_{ij}\left(1-\frac{y_{ij}}{z_j}\right)^{\beta}\right)^{\alpha/\beta}, \tag{3.25}$$

where α and β are nonnegative parameters, and it is used by Bourguignon and Chakravarty (1999, 2003) and Atkinson (2003) as a basis for demonstrating that the effect of an increasing association (correlation) rearrangement depends on whether the attributes are substitutes or complements, and that this corresponds to choosing $\alpha > \beta$ or $\alpha < \beta$.[37] Moreover, Atkinson (2003) demonstrates that the family of counting measures A_θ defined by (3.12) can be seen as a limiting case of (3.25) when α and β tend to zero with $\theta = \alpha/\beta$ and $w_1 = w_2 = 1$.

3.3.3.2 Partial Orderings

Most empirical studies consider few measures of poverty when ranking multidimensional distribution functions. A natural concern is that the conclusions reached in these studies are sensitive to the choice of the specific measures.[38] By drawing on Atkinson and Bourguignon (1982), Bourguignon and Chakravarty (2009) investigate what restrictions two alternative stochastic dominance conditions of the first-degree impose on the general family of two-dimensional poverty measures defined by

$$\Pi_p(H;z) = \int_0^{z_2}\int_0^{z_1} p(x_1,x_2;z_1,z_2)\mathrm{d}H(x_1,x_2), \tag{3.26}$$

where H is the bivariate distribution of the two attributes in question and $p(x_1, x_2; z_1, z_2)$ is the level of poverty associated with attribute levels (x_1, x_2) and poverty thresholds (z_1, z_2).[39] Let p_i' denote the derivative of p with respect to x_i and let p_{12}'' denote the second derivative of p with respect to x_1 and x_2. The following alternative presentation of a result from Bourguignon and Chakravarty (2009) provides three equivalent statements:

Theorem 3.2A

Let H and H^* be members of the family $\mathbf{H}$ of the bivariate distributions of the attributes (X_1, X_2) and let H_1 and H_1^* and H_2 and H_2^* be the associated marginal distributions of X_1 and X_2. Then the following statements are equivalent

[37] Brandolini (2009) and Madden (2011) use this index to study income and health poverty in selected European countries, and they analyze the sensitivity of results to different values of the parameters α and β. See also Bibi and El Lahga (2008).

[38] For similar concerns in the single-dimensional case, see Atkinson (1987), Zheng (1999), Spencer and Fisher (1992), Jenkins and Lambert (1997), and Aaberge and Atkinson (2013) who introduce poverty dominance criteria as a basis for obtaining more robust results.

[39] Poverty orderings of bivariate distributions are studied by Gravel and Moyes (2012), under the hypothesis that one attribute is cardinal and transferable between individuals, but the other is ordinal and non-transferable, and by Garcia-Diaz (2013), under the hypothesis of asymmetric treatment of the attributes proposed by Muller and Trannoy (2012).

(i) $H_i(x_i) \le H_i^*(x_i)$ for all $x_i < z_i, i = 1, 2$ *and* $H(x_1, x_2) \le H^*(x_1, x_2)$ for all $x_1 < z_1$ and $x_2 < z_2$

(ii) $\Pi_p(H; z) < \Pi_p(H^*; z)$ for all p where $p_i'(x_1, x_2) \le 0$ for $x_i < z_i, i = 1, 2$ *and* $p_{12}''(x_1, x_2) \ge 0$, for $x_1 < z_1$ and $x_2 < z_2$

(iii) *H^* can be obtained from H by sequences of Pigou–Dalton regressive transfers and/or a sequence of marginal-free (marginal distribution-preserving) correlation-increasing rearrangements.*

The equivalence between (i) and (ii) is proved by Bourguignon and Chakravarty (2009), and the equivalence between (ii) and (iii) is proved by Atkinson and Bourguignon (1982).

Theorem 3.2A shows that poverty measures Π_p that satisfy condition (ii) rank bivariate distributions according to first-degree stochastic dominance for attribute values below the poverty threshold in each dimension and first-degree two-dimensional stochastic dominance below both poverty thresholds. As indicated by Bourguignon and Chakravarty (2003, 2009), and demonstrated by Theorem 3.2A, the principle of correlation-increasing rearrangement (conditional on fixed marginal distributions) is associated with the intersection $H(x_1, x_2)$. Moreover, Theorem 3.2A shows that Π_p increases as a result of correlation-increasing rearrangement if the cross-derivative of p with respect to x_1 and x_2 is nonnegative. This is the reason why Bourguignon and Chakravarty, in this case, refer to the attributes as substitutes. The case in which the cross-derivative is negative corresponds to complements that are associated with nonincreasing poverty under a correlation-increasing rearrangement. An alternative presentation of this result is given by the following theorem:

Theorem 3.2B

Let H and H^* be members of the family $\boldsymbol{H}$ of the bivariate distributions of the attributes (X_1, X_2) and let H_1 and H_1^* and H_2 and H_2^* be the associated marginal distributions of X_1 and X_2. Then the following statements are equivalent

(i) $H_1(x_1) + H_2(x_2) - H(x_1, x_2) \le H_1^*(x_1) + H_2^*(x_2) - H^*(x_1, x_2)$ *for all* $x_1 < z_1$ *and/or* $x_2 < z_2$

(ii) $\Pi_p(H; z) < \Pi_p(H^*; z)$ *for all p where* $p_i'(x_1, x_2) \le 0$ for $x_i < z_i, i = 1, 2$ *and* $p_{12}''(x_1, x_2) \le 0$ for $x_1 < z_1$ and $x_2 < z_2$

(iii) *H^* can be obtained from H by sequences of Pigou–Dalton regressive transfers and/or a sequence of marginal-free (marginal distribution-preserving) correlation-decreasing rearrangements.*

For the proof of Theorem 3.2B, we refer to Atkinson and Bourguignon (1982) and Bourguignon and Chakravarty (2009). Note that Theorems 3.2A and 3.2B are, strictly speaking, only valid in cases for which the association (correlation) between X_1 and X_2 is positive. However, as demonstrated for the counting approach (see Theorems 3.1A and 3.1B), it is straightforward to extend Theorems 3.2A and 3.2B to cover the case in which the association (correlation) is negative.

In practice, multivariate distributions might often cross. This is shown, for instance, by Arndt et al. (2012) and Nanivazo (2014) in their multidimensional analyses of first order dominance for child poverty in Vietnam and Mozambique and in the Democratic Republic of Congo, respectively. Thus, it is helpful to introduce weaker criteria than first-degree dominance. Duclos et al. (2006a, 2007, 2008) consider second-degree and higher-degree dominance conditions when attributes are regarded as substitutes. A good example is when we consider poverty in relation to both income and wealth, which are perfect or very close substitutes. (Imperfect substitutability may derive from the lower degree of liquidity of an asset.) In these exercises, the asset poverty line is often defined with reference to the income poverty line, because it is taken to be the amount of wealth necessary to maintain the socially defined minimum standard for a certain period for someone who has no other economic sources (e.g., Brandolini et al., 2010; Haveman and Wolff, 2004). Bourguignon and Chakravarty (2009) remark that this approach gets close to the unidimensional approach because the poverty line of one attribute is assumed to be a function of the poverty line of the other attribute.

3.4. MULTIDIMENSIONAL INEQUALITY MEASUREMENT

The surge of research on the measurement of inequality in multiple dimensions is fairly recent, but the central question is far from new. In *Income Distribution, Value Judgments, and Welfare*, Fisher was not interested in money income, a "scalar," but in "real" income, that is, "a vector whose components are amounts of commodities" (Fisher, 1956, p. 382). His analysis was carried out by aggregating commodities either by using constant prices—to which he assigned "no particular significance . . . as market valuations of the commodities. Any arbitrary set of weights would do as well" (Fisher, 1956, p. 383, fn. 6)—or by means of individual utility functions. Social welfare was thus seen as an aggregation of individual preferences, in the tradition of what Sen (1977) has labeled "welfarism." The modern approach to measuring inequality in multiple dimensions generally departs from this identification by interpreting the individual utility function as "the observer's evaluation of the individual's welfare" (Kolm, 1977, p. 3), so that "the social criterion makes no use of information on individual *i*'s relative valuation of the different elements of [the vector of goods received by person *i*]" (Atkinson and Bourguignon, 1982, p. 184).

As in the study of a single variable pioneered by Kolm (1969) and Atkinson (1970), the analysis proceeds by investigating the conclusions that can be reached on the ranking of multivariate distributions by making alternative assumptions on the order of aggregation and the shape of the social welfare function, or on the desired properties of inequality indices. We first consider the extension of the Pigou–Dalton transfers principle. We then move to partial orderings and sequential dominance criteria, and lastly to inequality indices.[40]

[40] See Bradburd and Ross (1988) and Fluckiger and Silber (1994) for early proposals of multidimensional inequality indices.

3.4.1 Multidimensional Extensions of the Pigou–Dalton Transfers Principle

Some of the requirements typically specified in the univariate case can be directly transferred to the multivariate context. For instance, the requirement that the social evaluation pays no attention to any other individual characteristics than those included in the vector of attributes—the "anonymity principle"—does not pose any problem, and it may in fact be even less restrictive with multiple dimensions. On the other hand, the multidimensional extension of the Pigou–Dalton transfers principle is less straightforward. In its original formulation, it states that inequality should fall as income is transferred from a richer to a poorer person, without modifying their relative ranks. (The last condition is unnecessary if the anonymity principle is assumed.) There is no unique way to reformulate the principle when there are two or more dimensions (Kolm, 1977).[41]

A first possible generalisation is suggested by Fisher's approach discussed above. Suppose that there are r attributes and n individuals. The distribution is represented by the $n \times r$ matrix $\mathbf{X} = [x_{ij}]$, $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, r$, where x_{ij} is the quantity of attribute j enjoyed by individual i, and $x_i = (x_{i1}, x_{i2}, \ldots, x_{ir})$ is the vector of the attributes of individual i. If attributes are aggregated for each individual by a vector of weights (prices) $\mathbf{p}$, the comparison of two alternative distributions $\mathbf{X}$ and $\mathbf{Y}$ can be reduced to that of the two resulting univariate distributions $\mathbf{Xp}$ and $\mathbf{Yp}$: if $\mathbf{Yp}$ Lorenz-dominates $\mathbf{Xp}$ for any possible $\mathbf{p}$, then $\mathbf{Y}$ is socially preferable to $\mathbf{X}$. This dominance criterion is known as "price majorization," "budget majorization," or "directional majorization." The degree of appropriateness of market prices for attributes such as the health status represents a problem for this dominance criterion, and other reasons may lead one to question its ethical foundations (Trannoy, 2006).

A second possibility is then to conceive a Pigou–Dalton experiment in the multivariate context as a transfer simultaneously and identically involving all attributes. Suppose that there are two attributes and three individuals. A (strict) Pigou–Dalton (PD) transfer between individuals 1 and 2 can be defined as the transfer to the poorer individual of the fraction λ of the extra quantity of attribute j held by the richer individual, or $\lambda|x_{2j} - x_{1j}|$ with $0 < \lambda < 1$. Thus, the PD-transfer yields the new vectors $\mathbf{x}_1' = (1-\lambda)\mathbf{x}_1 + \lambda\mathbf{x}_2$ and $\mathbf{x}_2' = \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$, while $\mathbf{x}_3$ is unchanged. An example is shown in the left-top panel of Figure 3.4, where the empty diamonds correspond to the new distribution. This distribution is socially preferable to the original one, indicated by full diamonds, as it is obtained by a PD-transfer between 1 and 2 (there is no change for individual 3). More generally, a PD-transfer may be described by $\mathbf{Y} = \mathbf{TX}$ where $\mathbf{T}$ is the $n \times n$ matrix

[41] See also Das Gupta and Bhandari (1989), Dardanoni (1995), Fleurbaey and Trannoy (2003), Mosler (2004), Fleurbaey (2006), Savaglio (2006a,b), Diez et al. (2007), Nakamura (2012), and Banerjee (2014a,b).

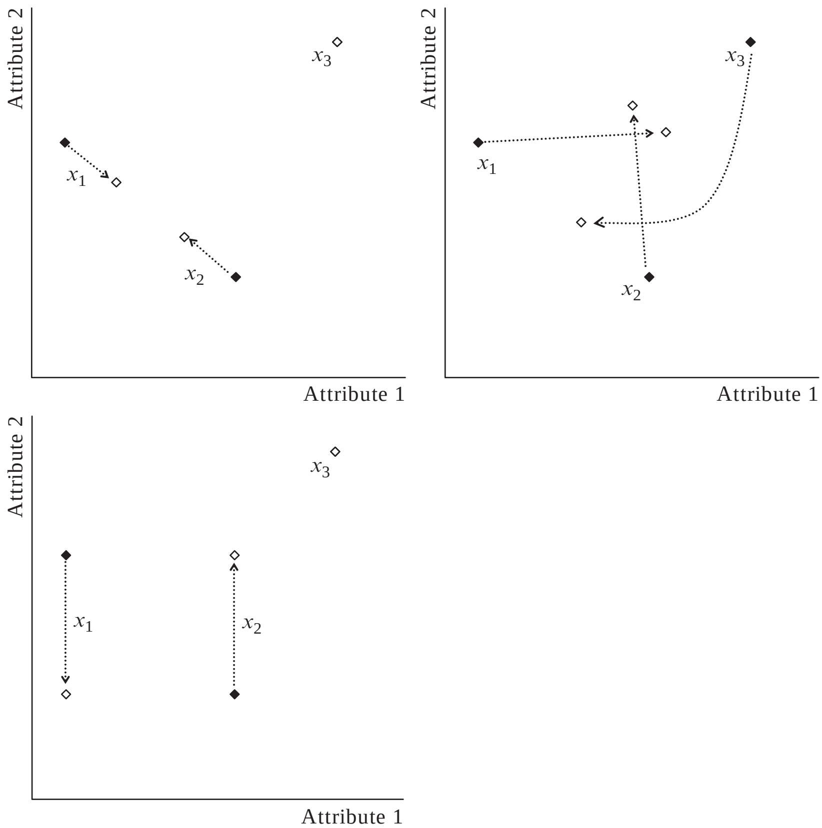


Figure 3.4 Examples of majorization criteria.

$\mathbf{T} = \lambda\mathbf{I} + (1-\lambda)\mathbf{\Pi}_{h,k}$ with $\mathbf{I}$ being the $n \times n$ identity matrix and $\mathbf{\Pi}_{h,k}$ being the $n \times n$ permutation matrix interchanging h and k (e.g., Weymark, 2006, p. 307). A distribution $\mathbf{Y}$ that is obtained from $\mathbf{X}$ by a sequence of PD-transfers is socially preferable to $\mathbf{X}$. This dominance criterion is known as "uniform Pigou-Dalton majorization", or "chain majorization" in Marshall and Olkin's (1979) terminology.

A sequence of PD-transfer matrices $\mathbf{T}$ yields a bistochastic matrix that is a nonnegative square matrix for which each row and each column sum up to 1. Though not all bistochastic matrices can be obtained by a sequence of PD-transfers, the multiplication of $\mathbf{X}$ by a bistochastic matrix is a form of averaging that makes the distribution less spread out. An alternative formulation of the dominance criterion is then to require that $\mathbf{Y}$ is socially preferable to $\mathbf{X}$ if there is a bistochastic $n \times n$ matrix $\mathbf{B}$ such that $\mathbf{Y} = \mathbf{BX}$ ("uniform majorization"). An example of a redistribution of this type, which cannot be obtained by a sequence of (strict) PD-transfers, is given in the right-top panel of Figure 3.4. It is clear, via visual examination, that the three individuals are closer each other after the averaging out performed by the **B** matrix; the deterioration suffered by individual 3 is socially acceptable by virtue of the anonymity principle.

There are two possible objections to these criteria. The first is that a change in one attribute does not affect the contribution to well-being of other attributes. We could however suppose that the correlation of attributes matter. Tsui (1999) introduces the concept of correlation-increasing transfer, which is an exchange of all attributes between two individuals after which one individual is left with the lowest endowment and the other with the maximum endowment of each attribute. By concentrating attributes, this type of transfer leads to a distribution that is less socially preferable than the original one. An example of "correlation-increasing majorization" is shown in the bottom-left panel of Figure 3.4. Figure 3.5 summarizes the majorization criteria.

The second objection is that, unlike income, many constituents of human welfare are not transferable. In general, it does not make much sense to talk of "transferring health" from a healthier individual to a sick one, with the possible exception of organ transplants (e.g. kidney and bone marrow). This has led Bosmans et al. (2009) to study the implications of formulating a version of the Pigou–Dalton principle that applies only to transferable attributes, and it has led Muller and Trannoy (2012) to examine dominance conditions under which attributes are asymmetric in the sense that one attribute (typically income) can be used to compensate for lower levels of other attributes (e.g., needs, health, etc.).

3.4.2 Partial Orderings and Sequential Dominance Criteria

As in the univariate case, conclusions based on summary measures of multidimensional inequality might be questioned. Thus, it is helpful to investigate their robustness by using partial orderings such as stochastic dominance criteria. The first-degree dominance criterion considered by Atkinson and Bourguignon (1982) was briefly discussed in Section 3.3.3.2. For a discussion of second-order multidimensional stochastic dominance and the conditions that this criterion imposes on the expected utility type of social welfare functions and associated measures of inequality, we refer to Atkinson and Bourguignon

Uniform Pigou–Dalton Majorization (UPD): $\mathbf{Y} \succ_{UPD} \mathbf{X}$ if and only if $\mathbf{Y} = \mathbf{TX}$ for some matrix $\mathbf{T}$ that is a finite product of PD transfer matrices and is not a permutation matrix.
Uniform Majorization (UM): $\mathbf{Y} \succ_{UM} \mathbf{X}$ whenever $\mathbf{Y} = \mathbf{BX}$, where $\mathbf{B}$ is a bistochastic matrix and $\mathbf{Y}$ cannot be derived by permuting the columns of $\mathbf{X}$.
Directional Majorization (DM): $\mathbf{Y} \succ_{DM} \mathbf{X}$ if and only if $\mathbf{Yp}$ strictly Lorenz-dominates $\mathbf{Xp}$ for any $\mathbf{p} \in R^m$.
Positive Directional Majorization (PDM): $\mathbf{Y} \succ_{PDM} \mathbf{X}$ if and only if $\mathbf{Yp}$ strictly Lorenz dominates $\mathbf{Xp}$ for any $\mathbf{p} \in R^m_{++}$.
Correlation-Increasing Majorization (CIM): $\mathbf{Y} \succ_{CIM} \mathbf{X}$ whenever $\mathbf{X}$ may be derived from $\mathbf{Y}$ by a permutation of columns and a finite sequence of correlation-increasing transfers at least one of which is strict.

Figure 3.5 Majorization criteria. *Source: adapted from Tsui (1999, pp. 149–152).*

(1982). Trannoy (2006) and Duclos et al. (2011) propose extensions of the results provided by Atkinson and Bourguignon (1982). Koshevoy (1995, 1998) and Koshevoy and Mosler (1996, 1997, 2007) introduce an alternative approach based on a multidimensional generalization of the Lorenz curve. Note that the equivalence between second-degree stochastic dominance and first-degree Lorenz dominance for fixed means does not hold in the multidimensional case.

The elaboration of sequential dominance criteria for the bivariate asymmetric space of income and household composition has been an early topic of the research on partial orderings in a multidimensional framework. Following Atkinson and Bourguignon (1987), many authors have seen the advantage of this approach over the standard income equivalization procedure in the fact that it only requires ranking family types in terms of needs, without specifying how much needier one family type is than another. Bourguignon (1989), Atkinson (1992), Jenkins and Lambert (1993), Moyes (2012), Chambaz and Maurin (1998), Ok and Lambert (1999), Ebert (2000), Lambert and Ramos (2002), Duclos and Makdissi (2005), Decoster and Ooghe (2006), and Zoli and Lambert (2012) belong to this branch of research, with a focus either on poverty or on inequality. Sequential dominance analysis can be applied to other bivariate distributions. Brandolini and D'Alessio (1998) present an early application to the joint distribution of equivalent income and health in Italy, whereas Duclos and Échevin (2011) and Madden (2014) carry out a similar exercise to compare Canada and the United States. Duclos et al. (2006b) study the joint distributions of household expenditure and children's heights in Ghana, Madagascar, and Uganda. Bérenger and Bresson (2012) use sequential dominance to test whether growth is "pro-poor" when poverty is measured by income and another discrete well-being attribute. Sequential dominance criteria for more than two attributes are presented by Gravel et al. (2009), Gravel and Mukhopadhyay (2010), and Muller and Trannoy (2011). McCaig and Yatchew (2007) and Batana and Duclos (2011) have developed statistical inference techniques to test dominance.

3.4.3 Measures of Multidimensional Inequality

As for the measurement of multidimensional deprivation and poverty, the informational basis defined by the order of aggregation plays a crucial role in measurement of multidimensional inequality as well. Thus, it is helpful to make a distinction between measures of multidimensional measures of inequality for which the order of aggregation either begins with aggregating across individuals for each single attribute or across attributes for each individual. In the former case, we obtain measures of overall inequality that aggregate inequality over each of the attributes. If we invert the order of aggregation, we derive an overall measure of inequality that aggregates synthetic functions of the attributes across individuals. The latter approach embeds the association between the achievements in the various dimensions into an overall indicator of individual achievements.

3.4.3.1 Two-Stage Approaches: First Aggregating Across Individuals

Two-stage approaches either aggregate individuals' achievements on each dimension and then the resulting attribute-specific indicators over the r dimensions, or they aggregate the single attributes into individual-specific well-being indicators, before aggregating these individual indicators into a summary measure of multidimensional inequality. The former approach forms the basis of the Inequality-adjusted Human Development Index (*IHDI*; e.g., UNDP, 2013), which belongs to the class of distribution-sensitive composite indices proposed by Foster et al. (2005), as well as of the following family of multidimensional generalized-Gini coefficients proposed by Gajdos and Weymark (2005):

$$J_{\tau w}(F) = 1 - \frac{W_{\tau w}(F)}{W_{\tau w}\left(F_{\text{equal}}\right)}, \tag{3.27}$$

where $W_{\tau w}(F)$ and $W_{\tau w}(F_{\text{equal}})$ are defined by

$$W_{\tau w}(F) = \begin{cases} \left[\sum_{j=1}^{r} \tau_j \left(\sum_{i=1}^{n} w_{ij} x_{ij}\right)^{\alpha}\right]^{1/\alpha} & \text{when}\, \alpha \neq 0 \\ \prod_{j=1}^{r} \left(\sum_{i=1}^{n} w_{ij} x_{ij}\right)^{\tau_j} & \text{when}\, \alpha = 0, \end{cases} \tag{3.28}$$

and

$$W_{\tau w}\left(F_{\text{equal}}\right) = \begin{cases} \left[\sum_{j=1}^{r} \tau_j \mu_j^{\alpha}\right]^{1/\alpha} & \text{when}\, \alpha \neq 0 \\ \prod_{j=1}^{r} \mu_j^{\tau_j} & \text{when}\, \alpha = 0, \end{cases} \tag{3.29}$$

where μ_j is the mean of attribute j.

Gajdos and Weymark (2005) demonstrate that the family of social evaluation functions $W_{\tau w}(F)$ is characterized by the following set of distributional associated axioms: *uniform Pigou–Dalton majorization principle* (*UPD*), *strong attribute separability* (*SAS*), *weak comonotonic additivity* (*WCA*), and *homotheticity* (*HOM*), as well as the conventional non-distributional axioms *ordering*, *continuity*, and *monotonicity*. *UPD* is a multidimensional Pigou–Dalton transfer principle. *SAS* requires that any subset of the attributes is independent of the other attributes. *WCA* is a multidimensional extension of the weak independence of income source axiom imposed by Weymark (1981) on the ordering of univariate income distributions, which is equivalent to the *dual independence* axiom discussed in Section 3.3. *HOM* is an extension of the scale invariance axiom for unidimensional inequality measures and requires that a common proportional change in the measurement units of the attributes should not affect the social evaluation ordering.[42]

By specifying $\alpha = 1$ and $\tau_j = 1/r$ in (3.28) and (3.29), $J_{\tau w}(F)$ becomes a weighted average of the attribute-specific generalized-Gini coefficients introduced by Donaldson and Weymark (1980). Alternatively, by choosing $\tau_j = 1/\mu_j$, $J_{\tau w}(F)$ becomes equal to the arithmetic mean of the attribute-specific generalized-Gini coefficients, previously proposed by Koshevoy and Mosler (1997).[43] Replacing *WCA* with a multidimensional extension of the *independence* axiom gives a normative justification of a multidimensional Atkinson family similar to the generalized-Gini family (3.27).

These types of multidimensional inequality measures ignore the impact of the association between attributes on overall inequality, and therefore, they do not exploit all information when individual-level data on multiple attributes are available.

3.4.3.2 Two-Stage Approaches: First Aggregating Across Attributes

Measures that capture the association between attributes can be derived either from a two-stage aggregation approach or from a direct one-stage approach. The two-stage approach originally proposed by Maasoumi (1986, 1989, 1999) uses a (common) utility-like function (measure of well-being) to aggregate the attributes for each individual in the first stage and a univariate inequality measure to aggregate the utility-like values across individuals in the second stage. Seth (2013) and Bosmans et al. (2013a) give the two-stage approach a normative justification. Let the social evaluation (or welfare) function W associated with the two-stage approach be defined by

$$W(F) = v(u(x_1), u(x_2), \ldots, u(x_n)), \tag{3.30}$$

where $x_i = (x_{i1}, x_{i2}, \ldots, x_{ir})$ is the attribute bundle of individual i, $i = 1, 2, \ldots, n$; F is the multidimensional distribution of the r attributes; and u is the common utility-like function. Bosmans et al. (2013a) demonstrate that $W(F)$ is characterized by the following axioms:[44] *monotonicity*, *continuity*, *normalization* (provides a cardinalization of the social evaluation function), *anonymity* (makes the utility function common to all individuals), *homotheticity* ($W(F)$ is invariant to a common proportional change in each attribute), *weak uniform majorization* (progressive transfers uniformly applied to each attribute do not

[42] See Gajdos and Weymark (2005) for a discussion of strengthening the scale invariance axiom to allow for independent proportional changes in the measurement units of the attributes, which is required when one considers both monetary and nonmonetary attributes.

[43] Okamoto (2009) provides a decomposition of this class of multivariate Gini indices which satisfies the *completely identical distribution* condition, whereby the between-group inequality is equal to zero if the distribution is the same within all population subgroups.

[44] Seth (2013) provides an axiomatic characterization of a two-stage approach in which generalized means form the basis of the aggregation in each stage. See also Lasso de la Vega et al. (2010) who consider the two-stage generalized mean approach for analyzing multidimensional deprivation distributions.

decrease $W(F)$), and *individualism* (social evaluation is made in two steps: the first step aggregates across attributes for each individual and the second step aggregates the aggregated attributes across individuals).

Thus, several of the proposed families of multidimensional inequality measures can be ethically justified by drawing on the characterization results of Bosmans et al. (2013a). For example, the common utility-like function can be specified as

$$u(x_i) = \sum_j w_j x_{ij}, \tag{3.31}$$

where w_j is the weight associated with attribute j, equal across individuals, and weights are normalized to sum to unity. The hypothesis of additive separability used in (3.31) rules out attributes that are not perfect substitutes. As suggested by Maasoumi (1986), a straightforward generalization of (3.31) is offered by the class of utility functions showing constant elasticity of substitution (CES)

$$u(x_i) = \begin{cases} \left[\sum_j w_j x_{ij}^{-\beta}\right]^{1/\beta} & \beta \neq 0 \\ \prod_j x_{ij}^{w_j} & \beta = 0, \end{cases} \tag{3.32}$$

where β is a parameter governing the degree of substitution between the attributes. As β goes to infinity, the attributes are perfect complements, whereas they are perfect substitutes for $\beta = -1$. To aggregate the distribution of $u(x_i)$'s, Maasoumi (1986) proposes using either the entropy family or the Atkinson family of inequality measures.[45] Alternatively, in the second aggregation stage, we can rely on the family of rank-dependent measures, which includes the generalized-Gini family. List (1999), Banerjee (2010), and Decancq and Lugo (2012) characterize multidimensional Gini indices that aggregate first across attributes and then across individuals.

Tsui (1995, 1999) follows the direct one-stage approach. Tsui (1995) generalizes to the multivariate context Kolm's (1969) and Atkinson's (1970) analysis in which inequality is identified with the social welfare loss (see Sen, 1978, 1992, for a critique of ethical inequality indices). After restricting the class of social evaluation functions to be continuous, strictly increasing, anonymous, strictly quasi-concave, separable, and scale-invariant, Tsui (1995) derives the two following multidimensional (relative) inequality indices:[46]

[45] For instance, in their applications of Maasoumi's approach, Nilsson (2010), Justino (2012) and Rohde and Guest (2013) use the Theil indices.

[46] Abul Naga and Geoffard (2006), Brambilla and Peluso (2010), and Croci Angelini and Michelangeli (2012) provide decompositions of this class of indices into the univariate inequality indices of the attributes and a residual term capturing their joint distributions. See also Kobus (2012) for a stronger definition of decomposition by attributes. Diez et al. (2008) derive unit-consistent multidimensional inequality indices. Gigliarano and Mosler (2009) construct multidimensional indices of polarization. Abul Naga (2010) derives the large sample distribution of a class of multidimensional inequality indices including the Tsui index.

$$I_1 = 1 - \left[\frac{1}{n}\sum_i \prod_j \left(\frac{x_{ij}}{\mu_j}\right)^{r_j}\right]^{1/\sum_k r_k} \tag{3.33a}$$

$$I_2 = 1 - \prod_i \left[\prod_j \left(\frac{x_{ij}}{\mu_j}\right)^{r_j/\sum_k r_k}\right]^{1/n}, \tag{3.33b}$$

where μ_j is the mean of attribute j over all persons and parameters r_j's must satisfy certain restrictions. The separability condition implies that the attributes can be aggregated for every person i into an indicator of well-being $u(x_i) = \Pi_j x_{ij}{}^{w_j}$, where $w_j = r_j/\Sigma_k r_k$ can be seen as a normalized weight on attribute j. By replacing ε for $\Sigma_k r_k$, (3.33a) and (3.33b) can be rewritten as

$$I = \begin{cases} 1 - \left[\frac{1}{n}\sum_i \left(\frac{u(x_i)}{u(\mu)}\right)^{1-\varepsilon}\right]^{1/1-\varepsilon} & \varepsilon \neq 1 \\ 1 - \prod_i \left(\frac{u(x_i)}{u(\mu)}\right)^{1/n} & \varepsilon = 1 \end{cases}, \tag{3.34}$$

where $u(\mu) = \Pi_j \mu_j^{w_j}$ is the "representative" well-being of the society, or, in other words, the well-being of a person showing the mean achievement for each attribute. The restrictions on r_j transfer to w_j and ε; in the bivariate case, it is sufficient that $\varepsilon > 0$ and $0 < w_1 = 1 - w_2 < 1$.

This reformulation has four advantages. Firs, it demonstrates that the family defined by (3.33a) and (3.33b) could also be justified by the two-stage approach. Second, it shows the close link of the Tsui multidimensional inequality measure with the Atkinson univariate index applied to the $u(x)_i$'s, from which it differs only for the replacement of *mean* well-being with *representative* well-being. This is indeed the appropriate normalization because "maximizing social welfare under the constraint of fixed total resources of attributes ... requires to give each individual the average available quantity of attributes" (Bourguignon, 1999, p. 478). This observation exposes a conceptual diversity between the direct one-stage approach and the two-stage approach: the first normalizes by the representative well-being $u(\mu) = \Pi_j \mu_j^{w_j}$, but the latter uses the mean well-being $(1/n)\Sigma_i \Pi_j x_{ij}^{w_j}$. (Of course, the two indices coincide in the univariate case.) Third, (3.34) brings out the role of ε, or $\Sigma_k r_k$ in the original formulation, as the parameter that governs the degree of concavity, and hence inequality aversion, of the social evaluation function. In the univariate income space, the range of economically sensible values for ε can be restricted on the basis of considerations on the preference for redistribution. A similar analysis has not been conducted in the multivariate space of well-being, but "there is not necessarily any reason to change our views about the value of [ε] simply because we have moved to a higher dimensionality" (Atkinson, 2003, p. 59).[47] Fourth, (3.34) shows that the Tsui index allows for different weightings of the attributes (through the w_j's), but it makes no allowance for a variation in the degree of substitution between the attributes: the Cobb–Douglas functional form of the underlying well-being indicator implies that the elasticity of substitution between two attributes is uniformly equal to unity. In the bivariate case, a straightforward generalization is represented by the index derived by Bourguignon (1999) by assuming a CES functional form for the indicator of well-being, which has the Tsui index as a special case (see Lugo, 2007). Tsui (1999) examines alternative axioms that lead to characterizing a class of multidimensional generalized entropy measures.

3.4.3.3 Indices for Binary Variables

If information is restricted to marginal distributions of zero/one variables, an overall measure of inequality is a function of the proportions of people with attribute values above each of the attribute-specific thresholds, which means that they do not suffer from deprivation in these dimensions.

By contrast, when multiple attributes are observed for the same individuals, let p_j be the proportion of people with j attributes that take values above the attribute-specific threshold levels, and $G(k) = \sum_{j=0}^{k} p_j$ can be the cumulative proportion of people with k or fewer attributes that take values above the attribute-specific threshold levels. Then, similar to the discussion for the distribution of deprivation counts in Section 3.3.1, the social evaluation function

$$W_{\Psi}(G) = r - \sum_{k=0}^{r-1} \Psi\left(\sum_{j=0}^{k} p_j\right) \tag{3.35}$$

yields the following measures of dual multidimensional inequality:

$$I_{\Psi}(G) = 1 - \frac{W_{\Psi}(G)}{\nu} = 1 - \frac{r - \sum_{k=0}^{r-1} \Psi\left(\sum_{j=0}^{k} p_j\right)}{\nu}, \tag{3.36}$$

where ν is the average number of individual achievements above the attribute thresholds, and Ψ, with $\Psi(0) = 0$ and $\Psi(1) = 1$, is a nonnegative and nondecreasing concave function capturing the preferences of a social evaluator who supports axioms similar to those underlining the rank-dependent utility theory of Yaari (1987).

[47] In the analysis of income inequality, Atkinson and Brandolini (2010) suggest that plausible values for ε are between 0.3 and 3. This range includes the values used by Lugo (2007) and Brandolini (2009) in their empirical analyses. In a cross-national comparison of multidimensional inequality, Aristei and Perugini (2010) use country-specific values of ε, ranging from 1.04 to 1.77, estimated from national tax structures.

Note that $G(k) = 1 - F(r - k - 1)$, where F and μ are the count distribution of deprivations and the mean number of deprivations discussed in Section 3.3.1, which means that $\nu + \mu = r$. In other words, the sum of the mean number of deprivations and the mean number of achievements is necessarily equal to the number of attributes. By specifying $\Psi(t) = 1 - \Gamma(1 - t)$ it can be demonstrated that this adding-up condition is also satisfied by the sum of the deprivation measure D and social evaluation function W,

$$\begin{aligned} W_{\Psi}(G) &= W_{1-\Gamma}(1 - F(r - k - 1)) \\ &= r - \sum_{k=0}^{r-1}\left(1 - \Gamma\left(1 - \sum_{j=0}^{r-k-1} p_j\right)\right) = \sum_{k=0}^{r-1}\left(\Gamma\left(\sum_{j=0}^{k} q_j\right)\right) = r - D_{\Gamma}(F). \end{aligned}$$

Thus, inequality in the count distribution of achievements, rather than deprivations, can be given the following alternative expression:

$$I_{\Psi}(G) = 1 - \frac{r - D_{\Gamma}(F)}{r - \mu} = \frac{D_{\Gamma}(F) - \mu}{r - \mu} = \frac{\Delta_{\Gamma}(F)}{r - \mu}, \tag{3.37}$$

where Γ is a nondecreasing convex function. Inequality in the distribution of achievements is equivalent to the relative spread of deprivations (divided by the difference between the mean number of deprivations and achievements). Note that the notion of inequality is closely associated with the intersection approach discussed in Section 3.3, whereas the union approach is in conflict with the notion of inequality.

The primal analogs to $W_{\psi}(G)$ and $I_{\Psi}(G)$, and the counterpart of $d_{\gamma}(F)$ defined by (3.10), are given by

$$w_{\xi}(G) = \sum_{k=0}^{r} \xi(k) p_k \tag{3.38}$$

and

$$J_{\xi}(G) = 1 - \frac{w_{\xi}(G)}{\xi(\nu)} = 1 - \frac{\sum_{k=0}^{r} \xi(k) p_k}{\xi(\nu)}, \tag{3.39}$$

where ξ is a nonnegative and nondecreasing concave function capturing the preferences of a social evaluator who supports the *independence* axiom for orderings defined on the set of G-distributions. By specifying $\xi(k) = \gamma(r) - \gamma(r - k)$ and inserting for $p_k = q_{r-k}$, we get

$$w_{\xi}(G) = \sum_{k=0}^{r} \xi(k) p_k = \sum_{k=0}^{r} \xi(k) q_{r-k} = \sum_{k=0}^{r} \xi(r-k) q_k = \sum_{k=0}^{r} (\gamma(r) \; - \gamma(k)) q_k \; = \gamma(r) - d_{\gamma}(F)$$

and $\xi(\nu) = \gamma(r) - \gamma(\mu)$, which yield the following alternative expression for J_{ξ}

$$J_\xi(G) = 1 - \frac{\gamma(r) - d_\gamma(F)}{\xi(\nu)} = \frac{d_\gamma(F) - \gamma(\mu)}{\gamma(r) - \gamma(\mu)} = \frac{\delta_\gamma(F)}{\gamma(r) - \gamma(\mu)}, \tag{3.40}$$

where $\delta_\gamma(F)$ is defined by (3.11) and γ is a nondecreasing convex function such that $\gamma(\mu) \leq d_\gamma(F) \leq \gamma(r)$.

3.5. SUMMARY AND CONCLUSIONS

Since the 1990s, the measurement of multidimensional inequality and poverty has turned into a thriving research area. Novel analytical results have accompanied a massive production of applied research. The increasing availability of new and rich databases has fueled the growth, but this process would have not been possible without the spreading of new conceptualizations of well-being, prominently the capability approach, and of a policy orientation more inclined to consider the nuances of human well-being. The progress has not always been coherent: applied researchers have sometimes moved from available data, unaware of analytical developments, and theoretical researchers have sometimes ignored the applicability of their results to real data. This is common when development is rapid, and it can contribute to explaining why we have enriched our toolbox with so many new instruments, but we still disagree on whether and how to use them. Our aim in this chapter has been to provide a manual to this toolbox, drawing connections between different strands of the literature, clarifying some ambiguities, and exposing the strict link between analytical tools and the characteristics of the data available for the analysis.

The informational basis of the analysis is indeed crucial: tools intended for cardinal or categorical variables need not be appropriate for dichotomous variables, which often represent the bulk of the available information. This is one reason why we have paid special attention to the counting approach. Another one is expository convenience: the role of marginal distributions and the association between the attributes are particularly transparent for dichotomous variables, especially in the two-dimensional case, although the descriptive and normative issues are similar to those of continuous variables. However, the main motivation for this choice has been the attempt to bridge the gap between a copious empirical literature and a still relatively underdeveloped analytical elaboration. We have derived dominance criteria and measures of deprivation by exploiting the fact that counting deprivations brings us back to a univariate space. Thus, the social evaluation of distributions of deprivation counts is in many respects analogous to the social evaluation of income distributions, although it implicitly accounts for the association among the deprivation indicators. Of course, concave preferences in the income space correspond to convex preferences in the space of deprivations counts, which represent "bads" (loss in welfare) rather than "goods" (gains in welfare). However, although convex preferences are ruled out in the analysis of income distributions because they would yield a social evaluation function violating the Pigou–Dalton principle of transfers, concave preferences are perfectly legitimate in the analysis of deprivation counts. This happens when we lean toward the union criterion, while convex preferences are associated with the intersection criterion. This example illustrates how the multidimensional case brings in new aspects that are unknown to the univariate case, but it also neatly exposes the strict connection between value judgements—where we draw the boundaries of poverty when there are multiple deprivations—and analytical tools—the degree of concavity/convexity of social preferences. There is clearly a need for further work on the analytical foundations of the social evaluation of distributions of deprivation scores.

The opposite situation characterizes the axiomatic treatment of poverty and inequality for continuous and categorical variables: a fairly rich theoretical apparatus does not appear to have yet made an impact on empirical investigations, except in sporadic applications. This may be due to the scarcity of suitable variables and databases, but it may also reflect the difficulty of discriminating among many equally sensible alternative tools. In addition to further developing and refining theoretical analysis, in this case, empirical work may play an important role in screening the most effective tools. Whatever the approach adopted, the quality and reliability of databases and the elaboration of inference tools, two aspects that we have virtually ignored in this chapter, are essential to supporting the validity of empirical analyses, especially when they are used to inform policy.

Yet, is it really worth devoting so much intellectual effort to develop the multidimensional analysis of poverty and inequality? It is an odd question at the end of such a long chapter, but as discussed in the introduction, the widely shared view that well-being, and hence poverty, is multidimensional does not necessarily imply that the social evaluation must itself be multidimensional. It may be for philosophical reasons or, more practically, because too much is lost in the process of aggregation. Once Sen (1987, p. 33) remarked that "the passion for aggregation makes good sense in many contexts, but it can be futile or pointless in others. . . . When we hear of variety, we need not invariably reach for our aggregator." On the other hand, the "eye-catching property" of the Human Development Index was praised by Streeten (1994, p. 235) as a powerful feature for its affirmation in the public debate, in spite of the theoretical weaknesses pointed out by its critics.[48] Three points may help us to find an answer to the question.

First, there is a pervasive demand by media commentators and policy-makers for multidimensional analyses. This demand must be met, not the least in order to avoid that such analyses are left to practitioners that conceive them as a bunching together of living standard indicators through some simple averaging or multivariate technique easily available in statistical and econometric packages. Empirical research confirms that broadening the evaluative space to include variables other than income can modify the picture drawn on

[48] For a recent example, see the exchange between Klugman et al. (2011a,b) and Ravallion (2011b, 2012a,b). See also Chakravarty (2011).

the basis of income alone. There is distinct informative value in adopting a multidimensional perspective. The theoretical work surveyed in this chapter facilitates the interpretation of empirical findings by bringing to the fore the implicit measurement assumptions and their economic meaning. If we estimate a lower deprivation index in the United Kingdom than in Italy using concave social preferences, as in Section 3.3.2.4, it is because we favor the union criterion, and hence, we tend to be relatively more worried by the spreading of a given number of deprivations across many people than by their concentration on fewer people who are hit more. If, on the contrary, we have convex preferences and are particularly concerned about those suffering from severe deprivations, we cannot unequivocally rank one country ahead of the other.

Second, the difficulties of multidimensional measurement should not be overstated. The choice of the degree of poverty or inequality aversion and the proper definition of indicators less familiar to us than income also arise in the univariate context. The problems that are new to the multivariate case are the weighting structure of the attributes and their degree of substitutability. Both these aspects are not technical hitches but rather the expression of implicit value judgements. Far from being a weakness of multidimensional approaches, the investigation of alternative assumptions is necessary to allow for the different views in the society. This is a sufficient reason for not devolving the resolution of these measurement problems to some statistical algorithm.

Third, the battery of instruments in our toolbox is ample. If we are reluctant to use a summary poverty or inequality index, we may fruitfully use sequential dominance analysis: it may yield a partial ordering, but it may sometimes be sufficient to evaluate, say, the impact on the distribution of well-being of alternative policies. The variety of our toolbox means that there is a middle ground between multidimensional summary indices and the dashboard approach, as stressed by Ferreira and Lugo (2013).

These are all good arguments in favor of multidimensional social evaluation. Are they also compelling enough to push us as far as to accept summary indices? Probably not, but two further comments are in order. The first is a pragmatic suggestion drawn from Bourguignon (1999, p. 483): when their building assumptions are properly understood, these indices can provide valuable insights if used "more as a dominance instrument than a strictly cardinal rule of comparison." The second is a somewhat deeper point. In a sense, the uneasiness with such a summary index in sectors of the economics profession may stem from the reluctance of those economists to abandon a utility-based conception of well-being. Only individuals are able to assess the trade-offs between the different constituents of well-being, and prices are the best available way to reveal such trade-offs, because they derive from the interactions of individuals in a market economy. If externalities, distortions, and missing markets prevent us from relying on prices as the aggregator of well-being dimensions, then the dashboard approach may be preferable, because no arbitrary weighting is imposed. The most developed conceptualization of multidimensional well-being to date, the capability approach, originates exactly from the rejection of a utility-based conception: "*valuing* a life and measuring the *happiness* generated in that life are two different exercises" (Sen, 1985, p. 12). If this is the founding aspect of multidimensional analysis, then the weighting of the different dimensions is an integral part of the evaluation exercise, and the reference to market prices loses much of its appeal. Social evaluation may attach more weight to work effort than that revealed by the wage, because jobs are characterized by other attributes that might contribute to reinforcing social integration. From this perspective, the practical solutions given to the selection of weights, which often boil down to equal weighting, may miss a decisive part of the evaluation. If this conjecture is correct, there is little chance that we will ever settle the controversy between the dashboard approach and summary indices.

ACKNOWLEDGMENTS

We are very grateful to Tony Atkinson and François Bourguignon for their inspiring discussions, insightful comments, and generous patience. We would also like to thank Sabina Alkire, Conchita D'Ambrosio, Jean-Yves Duclos, Stephen Jenkins, Eugenio Peluso, Luigi Federico Signorini, Henrik Sigstad and Claudio Zoli for their helpful comments. The views expressed here are solely ours; in particular, they do not necessarily reflect those of the Bank of Italy and Statistics Norway.

REFERENCES

Aaberge, R., 2000. Characterizations of Lorenz curves and income distributions. Soc. Choice Welf. 17, 639–653.
Aaberge, R., 2001. Axiomatic characterization of the Gini coefficient and Lorenz curve orderings. J. Econ. Theory 101, 115–132.
Aaberge, R., 2009. Ranking intersecting Lorenz curves. Soc. Choice Welf. 33, 235–259.
Aaberge, R., Atkinson, A.B., 2013. The Median as Watershed, Discussion Paper No. 749, Statistics Norway.
Aaberge, R., Brandolini, A., 2014. Social Evaluation of Multidimensional Count Distributions. Working Paper No. 2014-342, Society for the Study of Economic Inequality–ECINEQ, Verona.
Aaberge, R., Peluso, E., 2011. A Counting Approach for Measuring Multidimensional Deprivation, Working Papers no. 07/2011, Università di Verona, Dipartimento di Scienze economiche.
Abul Naga, R.H., 2010. Statistical inference for multidimensional inequality indices. Econ. Lett. 107, 49–51.
Abul Naga, R.H., Geoffard, P.Y., 2006. Decomposition of bivariate inequality indices by attributes. Econ. Lett. 90, 362–367.
Abul Naga, R.H., Yalcin, T., 2008. Inequality measurement for ordered response health data. J. Health Econ. 27, 1614–1625.
Alkire, S., 2002. Valuing Freedoms: Sen's Capability Approach and Poverty Reduction. Oxford University Press, Oxford.
Alkire, S., 2007. Choosing dimensions: the capability approach and multidimensional poverty. In: Kakwani, N., Silber, J. (Eds.), The Many Dimensions of Poverty. Palgrave Macmillan, Basingstoke, pp. 89–119.
Alkire, S., Foster, J., 2011a. Counting and multidimensional poverty measurement. J. Public Econ. 95, 476–487.
Alkire, S., Foster, J., 2011b. Understandings and misunderstandings of multidimensional poverty measurement. J. Econ. Inequal. 9, 289–314.
Alkire, S., Santos, M.E., 2010. Acute Multidimensional Poverty: A New Index for Developing Countries, United Nations Development Programme, Human Development Reports, Research Paper 2010/11, New York.

Alkire, S., Santos, M.E., 2013. A multidimensional approach: poverty measurement & beyond. Soc. Indic. Res. 112, 239–257.

Alkire, S., Santos, M.E., 2014. Measuring acute poverty in the developing world: robustness and scope of the multidimensional poverty index. World Dev. 59, 251–274.

Alkire, S., Foster, J., Santos, M.E., 2011. Where did identification go? J. Econ. Inequal. 9, 501–505.

Allison, R.A., Foster, J.E., 2004. Measuring health inequality using qualitative data. J. Health Econ. 23, 505–524.

Anand, S., Sen, A.K., 1997. Concepts of Human Development and Poverty: A Multi-Dimensional Perspective, United Nations Development Programme, Human Development Report 1997 Background Paper, New York.

Anderson, G., Crawford, I., Leicester, A., 2008. Efficiency analysis and the lower convex hull. In: Kakwani, N., Silber, J. (Eds.), Quantitative Approaches to Multidimensional Poverty Measurement. Palgrave-Macmillan, Basingstoke, pp. 176–191.

Aristei, D., Perugini, C., 2010. Preferences for redistribution and inequality in well-being across Europe. J. Policy Model. 32, 176–195.

Arndt, C., Distante, R., Hussain, M.A., Østerdal, L.P., Pham Lan, H., Ibraimo, M., 2012. Ordinal welfare comparisons with multiple discrete indicators: a first order dominance approach and application to child poverty. World Dev. 40, 2290–2301.

Asselin, L.-M., Anh, V.T., 2008. Multidimensional poverty and multiple correspondence analysis. In: Kakwani, N., Silber, J. (Eds.), Quantitative Approaches to Multidimensional Poverty Measurement. Palgrave-Macmillan, Basingstoke, pp. 80–103.

Atkinson, A.B., 1970. On the measurement of inequality. J. Econ. Theory 2, 244–263.

Atkinson, A.B., 1987. On the measurement of poverty. Econometrica 55, 749–764.

Atkinson, A.B., 1992. Measuring poverty and differences in family composition. Economica 59, 1–16.

Atkinson, A.B., 1998. Social exclusion, poverty, and unemployment. In: Atkinson, A.B., Hills, J. (Eds.), Exclusion, Employment and Opportunity. pp. 1–20, London School of Economics, Centre for Analysis of Social Exclusion, CASE paper 4.

Atkinson, A.B., 2003. Multidimensional deprivation: contrasting social welfare and counting approaches. J. Econ. Inequal. 1, 51–65.

Atkinson, A.B., Bourguignon, F., 1982. The comparison of multi-dimensioned distributions of economic status. Rev. Econ. Stud. 49, 183–201.

Atkinson, A.B., Bourguignon, F., 1987. Income distribution and differences in needs. In: Feiwel, G.R. (Ed.), Arrow and the Foundations of the Theory of Economic Policy. Macmillan, Basingstoke, pp. 350–370.

Atkinson, A.B., Bourguignon, F. (Eds.), 2000. In: Handbook of Income Distribution, vol. 1. North-Holland, Amsterdam.

Atkinson, A.B., Brandolini, A., 2010. On analyzing the world distribution of income. World Bank Econ. Rev. 24, 1–37.

Atkinson, A.B., Brandolini, A., forthcoming. Unveiling the ethics behind inequality measurement: Dalton's contribution to economics. Econ. J.

Atkinson, T., Cantillon, B., Marlier, E., Nolan, B., 2002. Social Indicators: The EU and Social Inclusion. Oxford University Press, Oxford.

Ayala, L., Jurado, A., Perez-Mayo, J., 2011. Income poverty and multidimensional deprivation: lessons from cross-regional analysis. Rev. Income Wealth 57, 40–60.

Banerjee, A.K., 2010. A multidimensional Gini index. Math. Soc. Sci. 60, 87–93.

Banerjee, A.K., 2014a. A multidimensional Lorenz dominance relation. Soc. Choice Welf. 42, 171–191.

Banerjee, A.K., 2014b. A multidimensional Lorenz dominance relation: some corrections. Soc. Choice Welf. http://dx.doi.org/10.1007/s00355-014-0808-9.

Basu, K., López-Calva, L.F., 2011. Functionings and capabilities. In: Arrow, K.J., Sen, A.K., Suzumura, K. (Eds.), Handbook of Social Choice and Welfare, vol. 2. Elsevier, Amsterdam, pp. 153–187.

Batana, Y.M., 2013. Multidimensional measurement of poverty among women in sub-Saharan Africa. Soc. Indic. Res. 112, 337–362.

Batana, Y.M., Duclos, J.-Y., 2011. Comparing multidimensional poverty with qualitative indicators of well-being. In: Deutsch, J., Silber, J. (Eds.), The Measurement of Individual Well-Being and Group Inequalities. Essays in Memory of Z.M. Berrebi. Routledge, Abingdon, pp. 280–297.
Battiston, D., Cruces, G., Lopez-Calva, L.F., Lugo, M.A., Santos, M.E., 2013. Income and beyond: multidimensional poverty in six Latin American Countries. Soc. Indic. Res. 112, 291–314.
Becker, G.S., Philipson, T.J., Soares, R.R., 2005. The quantity and quality of life and the evolution of World inequality. Am. Econ. Rev. 95, 277–291.
Belhadj, B., 2012. New weighing scheme for the dimensions in multidimensional poverty indices. Econ. Lett. 116, 304–307.
Belhadj, B., Limam, M., 2012. Unidimensional and multidimensional fuzzy poverty measures: new approach. Econ. Model. 29, 995–1002.
Bellani, L., 2013. Multidimensional indices of deprivation: the introduction of reference groups weights. J. Econ. Inequal. 11, 495–515.
Bellani, L., Hunter, G., Anand, P., 2013. Multidimensional welfare: do groups vary in their priorities and behaviours? Fisc. Stud. 34, 333–354.
Bennett, C.J., Mitra, S., 2013. Multidimensional poverty: measurement, estimation, and inference. Econ. Rev. 32, 57–83.
Bérenger, V., Bresson, F., 2012. On the "pro-poorness" of growth in a multidimensional context. Rev. Income Wealth 58, 457–480.
Betti, G., Verma, V., 2008. Fuzzy measures of the incidence of relative poverty and deprivation: a multidimensional perspective. JISS 17, 225–250.
Betti, G., D'Agostino, A., Neri, L., 2002. Panel regression models for measuring multidimensional poverty dynamics. JISS 11, 359–369.
Betti, G., Cheli, B., Lemmi, A., Verma, V., 2008. The fuzzy set approach to multidimensional poverty: the case of Italy in the 1990s. In: Kakwani, N., Silber, J. (Eds.), Quantitative Approaches to Multidimensional Poverty Measurement. Palgrave-Macmillan, Basingstoke, pp. 30–48.
Bibi, S., El Lahga, A.R., 2008. Robust ordinal comparisons of multidimensional poverty between South Africa and Egypt. Rev. Econ. Dev. 22, 37–65.
Birdsall, N., 2011. Comment on multi-dimensional indices. J. Econ. Inequal. 9, 489–491.
Bleichrodt, H., van Doorslaer, E., 2006. A welfare economics foundation for health inequality measurement. J. Health Econ. 25, 945–957.
Boarini, R., Mira D'Ercole, M., 2013. Going beyond GDP: an OECD perspective. Fisc. Stud. 34, 289–314.
Boland, P.J., Prochan, F., 1988. Multivariate arrangement increasing functions with applications in probability and statistics. J. Multivariate Anal. 25, 286–298.
Bosmans, K., Lauwers, L., Ooghe, E., 2009. A consistent multidimensional Pigou-Dalton transfer principle. J. Econ. Theory 144, 1358–1371.
Bosmans, K., Decancq, K., Ooghe, E., 2013a. What Do Normative Indices of Multidimensional Inequality Really Measure? CORE discussion paper 2013/35, Katholieke Universiteit Leuven.
Bosmans, K., Lauwers, L., Ooghe, E., 2013b. Prioritarian Poverty Comparisons with Cardinal and Ordinal Attributes, Center for Economic Studies discussion papers 13.10, Katholieke Universiteit Leuven.
Bossert, W., D'Ambrosio, C., Peragine, V., 2007. Deprivation and social exclusion. Economica 74, 777–803.
Bossert, W., Chakravarty, S.R., D'Ambrosio, C., 2013. Multidimensional poverty and material deprivation with discrete data. Rev. Income Wealth 59, 29–43.
Bourguignon, F., 1989. Family size and social utility. Income dominance criteria. J. Econ. 42, 67–80.
Bourguignon, F., 1999. Comment on 'multidimensioned approaches to welfare analysis' by E. Maasoumi. In: Silber, J. (Ed.), Handbook of Income Inequality Measurement. Kluwer, Boston, pp. 477–484.
Bourguignon, F., Chakravarty, S.R., 1999. A family of multidimensional poverty measures. In: Slottje, D.J. (Ed.), Advances in Econometrics, Income Distribution and Scientific Methodology. Essays in Honor of Camilo Dagum. Physica, Heidelberg, pp. 331–344.
Bourguignon, F., Chakravarty, S.R., 2003. The measurement of multidimensional poverty. J. Econ. Inequal. 1, 25–49.

Bourguignon, F., Chakravarty, S.R., 2009. Multi-dimensional poverty orderings. Theory and applications. In: Basu, K., Kanbur, R. (Eds.), Arguments for a Better World: Essays in Honor of Amartya Sen. In: Ethics, Welfare, and Measurement, vol. I. Oxford University Press, Oxford.

Bradburd, R.M., Ross, D.R., 1988. A general measure of multidimensional inequality. Oxf. Bull. Econ. Stat. 50, 429–433.

Brambilla, M.R., Peluso, E., 2010. A remark on 'decomposition of bivariate inequality indices by attributes' by Abul Naga and Geoffard. Econ. Lett. 108, 100.

Brandolini, A., 2009. On applying synthetic indices of multidimensional well-being: health and income inequalities in France, Germany, Italy, and the United Kingdom. In: Gotoh, R., Dumouchel, P. (Eds.), Against Injustice. The New Economics of Amartya Sen. Cambridge University Press, Cambridge, pp. 221–251.

Brandolini, A., D'Alessio, G., 1998. Measuring well-being in the functioning space. In: Chiappero Martinetti, E. (Ed.), Debating Global Society: Reach and Limits of the Capability Approach. Mimeo, Bank of Italy, Roma, pp. 91–156, Fondazione Giangiacomo Feltrinelli, Milano, 2009.

Brandolini, A., Magri, S., Smeeding, T.M., 2010. Asset-based measurement of poverty. J. Policy Anal. Manage. 29, 267–284.

Burchardt, T., Le Grand, J., Piachaud, D., 1999. Social exclusion in Britain 1991-1995. Soc. Policy Adm. 33, 227–244.

Callan, T., Nolan, B., 1991. Concepts of poverty and the poverty line. J. Econ. Surv. 5, 243–261.

Cao-Pinna, M., 1953. Le classi povere. In: Camera dei Deputati (Ed.), Atti della Commissione parlamentare di inchiesta sulla miseria in Italia e sui mezzi per combatterla. In: Indagini tecniche. Condizioni di vita delle classi misere, vol. II. Camera dei Deputati, Roma.

Cappellari, L., Jenkins, S.P., 2007. Summarising multiple deprivation indicators. In: Micklewright, J., Jenkins, S.P. (Eds.), Poverty and Inequality: New Directions. Oxford University Press, Oxford, pp. 166–184.

Caruso, G., Sosa-Escudero, W., Svarc, M., 2014. Deprivation and the dimensionality of welfare: a variable-selection cluster-analysis approach. Rev. Income Wealth. http://dx.doi.org/10.1111/roiw.12127.

Cavapozzi, D., Han, W., Miniaci, R., 2013. Alternative Weighting Structures for Multidimensional Poverty Assessment: Research Report 13018-EEF. University of Groningen.

Cerioli, A., Zani, S., 1990. A fuzzy approach to the measurement of poverty. In: Dagum, C., Zenga, M. (Eds.), Income and Wealth Distribution, Inequality and Poverty. Springer, Berlin, pp. 272–284.

Chakravarty, S.R., 2011. A reconsideration of the tradeoffs in the new human development index. J. Econ. Inequal. 9, 471–474.

Chakravarty, S.R., D'Ambrosio, C., 2006. The measurement of social exclusion. Rev. Income Wealth 52, 377–398.

Chakravarty, S.R., D'Ambrosio, C., 2013. A family of unit consistent multidimensional poverty indices. In: Bérenger, V., Bresson, F. (Eds.), Poverty and Social Exclusion Around the Mediterranean Sea. Springer, New York, pp. 75–88.

Chakravarty, S.R., Majumder, A., 2005. Measuring human poverty: a generalized index and an application using basic dimensions of life and some anthropometric indicators. J. Hum. Dev. 6, 275–299.

Chakravarty, S.R., Silber, J., 2008. Measuring multidimensional poverty: the axiomatic approach. In: Kakwani, N., Silber, J. (Eds.), Quantitative Approaches to Multidimensional Poverty Measurement. Palgrave-Macmillan, Basingstoke, pp. 192–209.

Chakravarty, S.R., Zoli, C., 2012. Stochastic dominance relations for integer variables. J. Econ. Theor. 147, 1331–1341.

Chakravarty, S.R., Mukherjee, D., Ranade, R.R., 1998. On the family of subgroup and factor decomposable measures of multidimensional poverty. In: Slottje, D. (Ed.), Research on Economic Inequality, vol. 8. JAI Press, Greenwich, pp. 175–194.

Chakravarty, S.R., Deutsch, J., Silber, J., 2008. On the Watts multidimensional poverty index and its decomposition. World Dev. 36, 1067–1077.

Coromaldi, M., Zoli, M., 2012. Deriving multidimensional poverty indicators: methodological issues and an empirical analysis for Italy. Soc. Indicat. Res. 107, 37–54.

Chambaz, C., Maurin, E., 1998. Atkinson and Bourguignon's dominance criteria: extended and applied to the measurement of poverty in France. Rev. Income Wealth 44, 497–513.
Cheli, B., 1995. Totally fuzzy and relative measures of poverty in dynamic context: an application to the British Household Panel Survey, 1991-1992. Metron 53, 115–134.
Cheli, B., Lemmi, A., 1995. A 'totally' fuzzy and relative approach to the multidimensional analysis of poverty. Econ. Notes 24, 115–134.
Cheli, B., Ghellini, G., Lemmi, A., Pannuzi, N., 1994. Measuring poverty in the countries in transition via TFR method: the case of Poland in 1990–1991. Stat. Trans. 1, 585–636.
Cherchye, L., Moesen, W., Van Puyenbroeck, T., 2004. Legitimately diverse, yet comparable: on synthesizing social inclusion performance in the EU. J. Common Mark. Stud. 42, 919–955.
Cherchye, L., Ooghe, E., Van Puyenbroeck, T., 2008. Robust human development rankings. J. Econ. Inequal. 6, 287–321.
Chiappero Martinetti, E., 1994. A new approach to evaluation of well-being and poverty by fuzzy set theory. G. Econ. Ann. Econ. 53 (n.s.), 367–388.
Chiappero Martinetti, E., 2000. A multidimensional assessment of well-being based on Sen's functioning approach. Riv. Int. Sci. Sociali 2, 207–239.
Cowell, F.A., 2000. Measurement of inequality. In: Atkinson, A.B., Bourguignon, F. (Eds.), Handbook of Income Distribution, pp. 87–166.
Cowell, F.A., Flachaire, E., 2012. Inequality with Ordinal Data. London School of Economics, London, and GREQAM, Marseille. Mimeo.
Croci Angelini, E., Michelangeli, A., 2012. Axiomatic measurement of multidimensional well-being inequality: some distributional questions. J. Socio Econ. 41, 548–557.
D'Ambrosio, C., Deutsch, J., Silber, J., 2011. Multidimensional approaches to poverty measurement: an empirical analysis of poverty in Belgium, France, Germany, Italy and Spain, based on the European panel. Appl. Econ. 43, 951–961.
Dagum, C., Costa, M., 2004. Analysis and measurement of poverty. Univariate and multivariate approaches and their policy implications. A case study: Italy. In: Dagum, C., Ferrari, G. (Eds.), Household Behaviour, Equivalence Scales, Welfare and Poverty. Physica, Heidelberg, pp. 221–271.
Danziger, S., van der Gaag, J., Taussig, M.K., Smolensky, E., 1984. The direct measurement of welfare levels: how much does it cost to make ends meet? Rev. Econ. Stat. 66, 500–505.
Dardanoni, V., 1995. On multidimensional inequality measurement. In: Dagum, C., Lemmi, A. (Eds.), Income Distribution, Social Welfare, Inequality, and Poverty, Research on Economic Inequality, vol. 6. JAI Press, Greenwich, pp. 201–207.
Das Gupta, S., Bhandari, S.K., 1989. Multivariate majorization. In: Gleser, L.J., Perlman, M.D., Press, S.J., Sampson, A.R. (Eds.), Contributions to Probability and Statistics. Essays in Honor of Ingram Olkin. Springer, New York, pp. 63–74.
Debreu, G., 1964. Continuity properties of Paretian utility. Int. Econ. Rev. 5, 285–293.
Decancq, K., 2014. Copula-based measurement of dependence between dimensions of well-being. Oxf. Econ. Pap. 66, 681–701.
Decancq, K., Lugo, M.A., 2012. Inequality of well-being: a multidimensional approach. Economica 79, 721–746.
Decancq, K., Lugo, M.A., 2013. Weights in multidimensional indices of wellbeing: an overview. Econ. Rev. 32, 7–34.
Decancq, K., Van Ootegem, L., Verhofstadt, E., 2013. What if we voted on the weights of a multidimensional well-being index? an illustration with Flemish data. Fisc. Stud. 34, 315–332.
Decancq, K., Fleurbaey, M., Maniquet, F., 2014. Multidimensional Poverty Measurement with Individual Preferences. Princeton University, William S. Dietrich II Economic Theory Center, Research Paper 058-2014.
Decoster, A., Ooghe, E., 2006. A bounded index test to make robust heterogeneous welfare comparisons. Rev. Income Wealth 52, 361–376.
Denny, K., 2002. New methods for comparing literacy across populations: insights from the measurement of poverty. J. R. Stat. Soc. 165, 481–493Part 3.

Desai, M., Shah, A., 1988. An econometric approach to the measurement of poverty. Oxf. Econ. Pap. 40, 505–522.
Deutsch, J., Silber, J., 2005. Measuring multidimensional poverty: an empirical comparison of various approaches. Rev. Income Wealth 51, 145–174.
Devicienti, F., Poggi, A., 2011. Poverty and social exclusion: two sides of the same coin or dynamically interrelated processes? Appl. Econ. 43, 3549–3571.
Dewilde, C., 2004. The multidimensional measurement of poverty in Belgium and Britain: a categorical approach. Soc. Indic. Res. 68, 331–369.
Di Tommaso, M.L., 2007. Children capabilities: a structural equation model for India. J. Socio Econ. 36, 436–450.
Diez, H., Lasso de la Vega, C., de Sarachu, A., Urrutia, A.M., 2007. A consistent multidimensional generalization of the Pigou-Dalton transfer principle: an analysis. B.E. J. Theor. Econ. 7,(Topics), art. 45.
Diez, H., Lasso de la Vega, C., Urrutia, A.M., 2008. Multidimensional unit- and subgroup-consistent inequality and poverty measures: some characterizations. In: Bishop, J., Zheng, B. (Eds.), Inequality and Opportunity: Papers from the Second ECINEQ Society Meeting, Research on Economic Inequality, vol. 16. Emerald Group Publishing Limited, Bingley, pp. 189–211.
Donaldson, D., Weymark, J.A., 1980. A single parameter generalization of the Gini indices of inequality. J. Econ. Theory 22, 67–86.
Duclos, J.-Y., Échevin, D., 2011. Health and income: a robust comparison of Canada and the US. J. Health Econ. 30, 293–302.
Duclos, J.-Y., Makdissi, P., 2005. Sequential stochastic dominance and the robustness of poverty orderings. Rev. Income Wealth 51, 63–88.
Duclos, J.-Y., Sahn, D.E., Younger, S.D., 2006a. Robust multidimensional poverty comparisons. Econ. J. 116, 943–968.
Duclos, J.-Y., Sahn, D.E., Younger, S.D., 2006b. Robust multidimensional spatial poverty comparisons in Ghana, Madagascar, and Uganda. World Bank Econ. Rev. 20, 91–113.
Duclos, J.-Y., Sahn, D.E., Younger, S.D., 2007. Robust multidimensional poverty comparisons with discrete indicators of well-being. In: Micklewright, J., Jenkins, S.P. (Eds.), Poverty and Inequality: New Directions. Oxford University Press, Oxford, pp. 185–206.
Duclos, J.-Y., Sahn, D.E., Younger, S.D., 2008. Using an Ordinal Approach to Multidimensional Poverty Analysis. In: Kakwani, N., Silber, J. (Eds.), Quantitative Approaches to Multidimensional Poverty Measurement. Palgrave-Macmillan, Basingstoke, pp. 244–261.
Duclos, J.-Y., Sahn, D.E., Younger, S.D., 2011. Partial multidimensional inequality orderings. J. Publ. Econ. 95, 225–238.
Duclos, J.-Y., Tiberti, L., forthcoming. Multidimensional poverty indices: a critical assessment. In: Adler, M., Fleurbaey, M. (Eds.), Oxford Handbook of Well-Being and Public Policy. Oxford University Press, Oxford.
Dutta, I., Foster, J., 2013. Inequality of happiness in the U.S.: 1972–2010. Rev. Income Wealth 59, 393–415.
Dutta, I., Pattanaik, P.K., Xu, Y., 2003. On measuring deprivation and the standard of living in a multidimensional framework on the basis of aggregate data. Economica 70, 197–221.
Ebert, U., 2000. Sequential generalized Lorenz dominance and transfer principles. Bull. Econ. Res. 52, 113–122.
Ebert, U., 2010. The decomposition of inequality reconsidered: weakly decomposable measures. Math. Soc. Sci. 60, 94–103.
Epstein, L., Tanny, S.M., 1980. Increasing generalized correlation: a definition and some economic consequences. Can. J. Econ. 13, 16–34.
Erikson, R., 1993. Description of inequality: the Swedish approach to welfare research. In: Nussbaum, M. C., Sen, A.K. (Eds.), The Quality of Life. Clarendon Press, Oxford, pp. 66–83.
Erikson, R., Goldthorpe, J.H., 1993. The Constant Flux. A Study of Class Mobility in Industrial Societies. Clarendon Press, Oxford.
Erikson, R., Uusitalo, H., 1986-87. The Scandinavian approach to welfare research. Int. J. Sociol. 16, 177–193.

Esposito, L., Chiappero Martinetti, E., 2010. Multidimensional poverty: restricted and unrestricted hierarchy among poverty dimensions. J. Appl. Econ. 13, 181–204.
European Commission, 2010. Europe 2020: A strategy for Smart, Sustainable and Inclusive Growth, COM (2010) 2020. final, Brussels, 3 March 2010.
Eurostat, 2014. Material Deprivation Rate—Economic Strain and Durables Dimension (Source: SILC) (ilc_sip8). http://epp.eurostat.ec.europa.eu/portal/page/portal/income_social_inclusion_living_conditions/data/database.
Federman, M., Garner, T.I., Short, K., Cutter IV, W.B., Kiely, J., Levine, D., McGough, D., McMillen, M., 1996. What does it mean to be poor in America? Mon. Labor Rev. 119 (5), 3–17.
Ferreira, F.H.G., 2011. Poverty is multidimensional. But what are we going to do about it? J. Econ. Inequal. 9, 493–495.
Ferreira, F.H.G., Lugo, M.A., 2013. Multidimensional poverty analysis: looking for a middle ground. World Bank Res. Obs. 28, 220–235.
Ferro Luzzi, G., Flückiger, Y., Weber, S., 2008. Multidimensional poverty: factor and cluster analysis. In: Kakwani, N., Silber, J. (Eds.), Quantitative Approaches to Multidimensional Poverty Measurement. Palgrave-Macmillan, Basingstoke, pp. 63–79.
Figari, F., 2012. Cross-national differences in determinants of multiple deprivation in Europe. J. Econ. Inequal. 10, 397–418.
Fisher, F.M., 1956. Income distribution, value judgments, and welfare. Q. J. Econ. 70, 380–424.
Fleurbaey, M., 2006. Social welfare, priority to the worst-off and the dimensions of individual well-being. In: Farina, F., Savaglio, E. (Eds.), Inequality and Economic Integration. Routledge, London, pp. 222–263.
Fleurbaey, M., Gaulier, G., 2009. International comparisons of living standards by equivalent incomes. Scand. J. Econ. 111, 597–624.
Fleurbaey, M., Trannoy, A., 2003. The impossibility of a Paretian egalitarian. Soc. Choice Welf. 21, 243–263.
Fluckiger, Y., Silber, J., 1994. The Gini index and the measurement of multidimensional inequality. Oxf. Bull. Econ. Stat. 56, 225–228.
Foster, J.E., Sen, A.K., 1997. On Economic Inequality after a Quarter Century. In: Sen, A.K. (Ed.), On Economic Inequality. Clarendon Press, Oxford, pp. 107–219 Expanded edition with a substantial annex by J.E. Foster and A.K. Sen.
Foster, J.E., Greer, J., Thorbecke, E., 1984. A class of decomposable poverty measures. Econometrica 52, 761–766.
Foster, J.E., López-Calva, L.F., Székely, M., 2005. Measuring the distribution of human development: methodology and an application to Mexico. J. Hum. Dev. 6, 5–29.
Fusco, A., Dickes, P., 2008. The Rasch model and multidimensional poverty measurement. In: Kakwani, N., Silber, J. (Eds.), Quantitative Approaches to Multidimensional Poverty Measurement. Palgrave-Macmillan, Basingstoke, pp. 49–62.
Fusco, A., Guio, A.-C., Marlier, E., 2010. Characterising the income poor and the materially deprived in European countries. In: Atkinson, A.B., Marlier, E. (Eds.), Income and Living Conditions in Europe. Publications Office of the European Union, Luxembourg, pp. 133–153.
Gajdos, T., Weymark, J.A., 2005. Multidimensional generalized Gini indices. Economic Theory 26, 471–496.
Garcia-Diaz, R., 2013. Poverty orderings with asymmetric attributes. B.E. J. Theor. Econ. 13, 347–361.
Gigliarano, C., Mosler, K., 2009. Constructing indices of multivariate polarization. J. Econ. Inequal. 7, 435–460.
Goedhart, T., Halberstadt, V., Kapteyn, A., van Praag, B.M.S., 1977. The poverty line: concept and measurement. J. Hum. Resour. 12, 503–520.
Gravel, N., Moyes, P., 2012. Ethically robust comparisons of bidimensional distributions with an ordinal attribute. J. Econ. Theory 147, 1384–1426.
Gravel, N., Mukhopadhyay, A., 2010. Is India better off today than 15 years ago? A robust multidimensional answer. J. Econ. Inequal. 8, 173–195.
Gravel, N., Moyes, P., Tarroux, B., 2009. Robust international comparisons of distributions of disposable income and regional public goods. Economica 76, 432–461.

Guio, A.C., 2005. Material deprivation in the EU. Eurostat, Statistics in Focus, Population and Social Conditions 21. Office for Official Publications of the European Communities, Luxembourg.
Halleröd, B., Larsson, D., Gordon, D., Ritakallio, V.-M., 2006. Relative deprivation: a comparative analysis of Britain, Finland and Sweden. J. Eur. Soc. Policy 16, 328–345.
Haveman, R., Wolff, E.N., 2004. The concept and measurement of asset poverty: levels, trends and composition for the U.S., 1983-2001. J. Econ. Inequal. 2, 145–169.
Hirschberg, J.G., Maasoumi, E., Slottje, D.J., 1991. Cluster analysis for measuring welfare and quality of life across countries. J. Econ. 50, 131–150.
Jenkins, S.P., Lambert, P.J., 1993. Ranking income distributions when needs differ. Rev. Income Wealth 39, 337–356.
Jenkins, S.P., Lambert, P.J., 1997. Three 'I's of poverty curves, with an analysis of UK poverty trends. Oxf. Econ. Pap. 49, 317–327.
Jorgenson, D.W., Slesnick, D.T., 1984a. Inequality in the distribution of individual welfare. In: Basmann, R.L., Rhodes, G.F. (Eds.), Advances in Econometrics, vol. 3. JAI Press, Greenwich, pp. 67–130.
Jorgenson, D.W., Slesnick, D.T., 1984b. Aggregate consumer behaviour and the measurement of inequality. Rev. Econ. Stud. 51, 369–392.
Jurado, A., Pérez-Mayo, J., 2012. Construction and evolution of a multidimensional well-being index for the Spanish regions. Soc. Indic. Res. 107, 259–279.
Justino, P., 2012. Multidimensional welfare distributions: empirical application to household panel data from Vietnam. Appl. Econ. 44, 3391–3405.
Kalmijn, W., Veenhoven, R., 2005. Measuring inequality of happiness in nations: in search for proper statistics. J. Happiness Stud. 6, 357–396.
Khan, A., Saboor, A., Ahmad, S., Ali, I., 2011. Mapping and measuring of multidimensional poverty in Pakistan: empirical investigations. Pak. J. Life Soc. Sci. 9, 121–127.
Kim, S.-G., 2014. Fuzzy multidimensional poverty measurement: an analysis of statistical behaviors. Soc. Indic. Res. http://dx.doi.org/10.1007/s11205-014-0616-8.
Klasen, S., 2000. Measuring poverty and deprivation in South Africa. Rev. Income Wealth 46, 33–58.
Klugman, J., Rodriguez, F., Choi, H.-J., 2011a. The HDI 2010: new controversies, old critiques. J. Econ. Inequal. 9, 249–288.
Klugman, J., Rodriguez, F., Choi, H.-J., 2011b. Response to Martin Ravallion. J. Econ. Inequal. 9, 497–499.
Kobus, M., 2012. Attribute decomposition of multidimensional inequality indices. Econ. Lett. 117, 189–191.
Kolm, S.-C., 1969. The optimal production of social justice. In: Margolis, J., Guitton, H. (Eds.), Public Economics. An Analysis of Public Production and Consumption and Their Relations to the Private Sectors. Macmillan, London, pp. 145–200.
Kolm, S.-C., 1977. Multidimensional egalitarianisms. Q. J. Econ. 91, 1–13.
Koshevoy, G.A., 1995. Multivariate Lorenz majorization. Soc. Choice Welf. 12, 93–102.
Koshevoy, G.A., 1998. The Lorenz zonotope and multivariate majorizations. Soc. Choice Welf. 15, 1–14.
Koshevoy, G.A., Mosler, K., 1996. The Lorenz zonoid of a multivariate distribution. J. Am. Stat. Assoc. 91, 873–882.
Koshevoy, G.A., Mosler, K., 1997. Multivariate Gini indices. J. Multivar. Anal. 60, 252–276.
Koshevoy, G.A., Mosler, K., 2007. Multivariate Lorenz dominance based on zonoids. AStA Adv. Stat. Anal. 91, 57–76.
Krishnakumar, J., 2008. Multidimensional measures of poverty and well-being based on latent variable models. In: Kakwani, N., Silber, J. (Eds.), Quantitative Approaches to Multidimensional Poverty Measurement. Palgrave-Macmillan, Basingstoke, pp. 118–134.
Krishnakumar, J., Ballon, P., 2008. Estimating basic capabilities: a structural equation model applied to Bolivia. World Dev. 36, 992–1010.
Krishnakumar, J., Nagar, A.L., 2008. On exact statistical properties of multidimensional indices based on principal components, factor analysis, MIMIC and structural equation models. Soc. Indic. Res. 86, 481–496.
Kuklys, W., 2005. Amartya Sen's Capability Approach. Springer, Berlin.

Lambert, P.J., Ramos, X., 2002. Welfare comparisons: sequential procedures for heterogenous populations. Economica 69, 549–562.

Lasso de la Vega, C., 2010. Counting poverty orderings and deprivation curves. In: Bishop, J.A. (Ed.), Studies in Applied Welfare Analysis: Papers from the Third ECINEQ Meeting, Research on Economic Inequality, vol. 18. Emerald Group Publishing Limited, Bingley, pp. 153–172.

Lasso de la Vega, C., Urrutia, A.M., 2011. Characterizing how to aggregate the individuals' deprivations in a multidimensional framework. J. Econ. Inequal. 9, 183–194.

Lasso de la Vega, C., Urrutia, A.M., De Sarachu, A., 2010. Characterizing multidimensional inequality measures which fulfil the Pigou-Dalton bundle principle. Soc. Choice Welf. 35, 319–329.

Le Breton, M., Peluso, E., 2010. Smooth inequality measurement: approximation theorems. J. Math. Econ. 46, 405–415.

Lelli, S., 2005. Using functionings to estimate equivalence scales. Rev. Income Wealth 51, 255–284.

List, C., 1999. Multidimensional Inequality Measurement: A Proposal. Nuffield College Working Paper in Economics, Nuffield College, Oxford.

Lovell, C.A.K., Richardson, S., Travers, P., Wood, L., 1994. Resources and functionings: a new view of inequality in Australia. In: Eichhorn, W. (Ed.), Models and Measurement of Welfare and Inequality. Springer, Heidelberg, pp. 787–807.

Lucchini, M., Assi, J., 2013. Mapping patterns of multiple deprivation and well-being using self-organizing maps: an application to Swiss household panel data. Soc. Indic. Res. 112, 129–149.

Lugo, M.A., 2007. Comparing multidimensional indices of inequality: methods and application. In: Bishop, J., Amiel, Y. (Eds.), Inequality and Poverty: Papers from the Society for the Study of Economic Inequality's Inaugural Meeting, Research on Economic Inequality, vol. 14. Elsevier JAI, Amsterdam, pp. 213–236.

Lustig, N., 2011. Multidimensional indices of achievements and poverty: What do we gain and what do we lose? An introduction to JOEI forum on multidimensional poverty. J. Econ. Inequal. 9, 227–234.

Maasoumi, E., 1986. The measurement and decomposition of multi-dimensional inequality. Econometrica 54, 991–997.

Maasoumi, E., 1989. Continuously distributed attributes and measures of multivariate inequality. J. Econ. 42, 131–144.

Maasoumi, E., 1999. Multidimensioned approaches to welfare analysis. In: Silber, J. (Ed.), Handbook of Income Inequality Measurement. Kluwer, Boston, pp. 437–477.

Maasoumi, E., Lugo, M.A., 2008. The information basis of multivariate poverty assessments. In: Kakwani, N., Silber, J. (Eds.), Quantitative Approaches to Multidimensional Poverty Measurement. Palgrave-Macmillan, Basingstoke, pp. 1–29.

Maasoumi, E., Nickelsburg, G., 1988. Multivariate measures of well-being and an analysis of inequality in the Michigan data. J. Bus. Econ. Stat. 6, 327–334.

Mack, J., Lansley, S., 1985. Poor Britain. Allen and Unwin, London.

Madden, D., 2011. Health and income poverty in Ireland, 2003-2006. J. Econ. Inequal. 9, 23–33.

Madden, D., 2014. Health and wealth on the roller-coaster: Ireland, 2003–2011. Soc. Indicat. Res. http://dx.doi.org/10.1007/s11205-014-0644-4.

Maquet, I., Stanton, D., 2012. Income indicators for the EU's social inclusion strategy. In: Besharov, D.J., Couch, K.A. (Eds.), Counting the Poor. New Thinking about European Poverty Measures and Lessons for the United States. Oxford University Press, New York, pp. 59–77.

Marlier, E., Atkinson, A.B., 2010. Indicators of poverty and social exclusion in a global context. J. Policy Anal. Manage. 29, 285–304.

Marlier, E., Atkinson, A.B., Cantillon, B., Nolan, B., 2007. The EU and Social Inclusion: Facing the Challenges. Policy Press, Bristol.

Marlier, E., Cantillon, B., Nolan, B., Van den Bosch, K., Van Rie, T., 2012. Developing and learning from EU measures of social inclusion. In: Besharov, D.J., Couch, K.A. (Eds.), Counting the Poor. New Thinking about European Poverty Measures and Lessons for the United States. Oxford University Press, New York, pp. 299–341.

Marshall, A.W., Olkin, I., 1979. Inequalities: Theory of Majorization and Its Applications. Academic Press, New York.

Mayer, S.E., Jencks, C., 1989. Poverty and the distribution of material hardship. J. Hum. Resour. 21, 88–113.
McCaig, B., Yatchew, A., 2007. International welfare comparisons and nonparametric testing of multivariate stochastic dominance. J. Appl. Econometrics 22, 951–969.
Merz, J., Rathjen, T., 2014a. Multidimensional time and income poverty: well-being gap and minimum 2DGAP poverty intensity—German evidence. J. Econ. Inequal. http://dx.doi.org/10.1007/s10888-013-9271-6.
Merz, J., Rathjen, T., 2014b. Time and income poverty: an interdependent multidimensional poverty approach with German time use diary data. Rev. Income Wealth 60, 450–479.
Micklewright, J., Schnepf, S.V., 2007. Inequality of learning in industrialized countries. In: Micklewright, J., Jenkins, S.P. (Eds.), Poverty and Inequality: New Directions. Oxford University Press, Oxford, pp. 129–145.
Mitra, S., Jones, K., Vick, B., Brown, D., McGinn, E., Alexander, M.J., 2013. Implementing a multidimensional poverty measure using mixed methods and a participatory framework. Soc. Indicat. Res. 110, 1061–1081.
Mohanty, S.K., 2011. Multidimensional poverty and child survival in India. PLoS One 6, e26857.
Mosler, K., 2004. Restricted Lorenz dominance of economic inequality in one and many dimensions. J. Econ. Inequal. 2, 89–103.
Moyes, P., 2012. Comparisons of heterogenous distributions and dominance criteria. J. Econ. Theory 147, 1351–1383.
Muller, C., Trannoy, A., 2011. A dominance approach to the appraisal of the distribution of well-being across countries. J. Public Econ. 95, 239–246.
Muller, C., Trannoy, A., 2012. Multidimensional inequality comparisons: a compensation perspective. J. Econ. Theory 147, 1427–1449.
Nakamura, K., 2012. Extension of Generalized Lorenz Dominance Criterion to Multivariate Attributes. Faculty of Economics, University of Toyama, Working Paper 266.
Nanivazo, M., 2014. First order dominance analysis: child wellbeing in the democratic republic of Congo. Soc. Indicat. Res. http://dx.doi.org/10.1007/s11205-014-0673-z.
Navarro, C., Ayala, L., 2008. Multidimensional housing deprivation indices with application to Spain. Appl. Econ. 40, 597–611.
Nelsen, R.B., 1998. An Introduction to Copulas. Lecture Notes in Statistics. Springer, Heidelberg.
Nilsson, T., 2010. Health, wealth and wisdom: exploring multidimensional inequality in a developing country. Soc. Indicat. Res. 95, 299–323.
Nolan, B., Whelan, C.T., 1996a. Resources, Deprivation, and Poverty. Clarendon Press, Oxford.
Nolan, B., Whelan, C.T., 1996b. The relationship between income and deprivation: a dynamic perspective. Rev. Econ. 3, 709–717.
Nolan, B., Whelan, C.T., 2007. On the multidimensionality of poverty and social exclusion. In: Micklewright, J., Jenkins, S.P. (Eds.), Poverty and Inequality: New Directions. Oxford University Press, Oxford, pp. 146–165.
Nolan, B., Whelan, C.T., 2010. Using non-monetary deprivation indicators to analyze poverty and social exclusion: Lessons from Europe? J. Policy Anal. Manage. 29, 305–325.
Nolan, B., Whelan, C.T., 2011. Poverty and Deprivation in Europe. Oxford University Press, Oxford.
Nussbaum, M.C., 1990. Aristotelian social democracy. In: Douglass, R.B., Mara, G., Richardson, H. (Eds.), Liberalism and the Good. Routledge, New York, pp. 203–252.
Nussbaum, M.C., 1993. Non-relative virtues: an Aristotelian approach. In: Nussbaum, M.C., Sen, A.K. (Eds.), The Quality of Life. Clarendon Press, Oxford, pp. 242–269.
Nussbaum, M.C., 2003. Capabilities as fundamental entitlements: Sen and social justice. Fem. Econ. 9, 33–59.
Ok, E., Lambert, P., 1999. On evaluating social welfare by sequential generalized Lorenz dominance. Econ. Lett. 63, 45–53.
Okamoto, M., 2009. Decomposition of Gini and multivariate Gini indices. J. Econ. Inequal. 7, 153–177.
Pattanaik, P., Reddy, S.G., Xu, Y., 2011. On measuring deprivation and living standards of societies in a multi-attribute framework. Oxf. Econ. Pap. 64, 43–56.

Peichl, A., Pestel, N., 2013a. Multidimensional affluence: theory and applications to Germany and the US. Appl. Econ. 45, 4591–4601.
Peichl, A., Pestel, N., 2013b. Multidimensional well-being at the top: evidence for Germany. Fisc. Stud. 34, 355–371.
Pérez-Mayo, J., 2005. Identifying deprivation profiles in Spain: a new approach. Appl. Econ. 37, 943–955.
Pérez-Mayo, J., 2007. Latent vs. fuzzy methodology in multidimensional poverty analysis. In: Bishop, J., Amiel, Y. (Eds.), Inequality and Poverty: Papers from the Society for the Study of Economic Inequality's Inaugural Meeting, Research on Economic Inequality, vol. 14. Elsevier JAI, Amsterdam, pp. 95–117.
Permanyer, I., 2014. Assessing individuals' deprivation in a multidimensional framework. J. Dev. Econ. 109, 1–16.
Pisati, M., Whelan, C.T., Lucchini, M., Maître, B., 2010. Mapping patterns of multiple deprivation using self-organising maps: an application to EU SILC data for Ireland. Soc. Sci. Res. 39, 405–418.
Poggi, A., 2007a. Does persistence of social exclusion exist in Spain? J. Econ. Inequal. 5, 53–72.
Poggi, A., 2007b. Social exclusion mobility in Spain, 1994–2001. In: Bishop, J., Amiel, Y. (Eds.), Inequality and Poverty: Papers from the Society for the Study of Economic Inequality's Inaugural Meeting, Research on Economic Inequality, vol. 14. Elsevier JAI, Amsterdam, pp. 71–94.
Poggi, A., Ramos, X., 2011. Empirical modeling of deprivation contagion among social exclusion dimensions. In: Deutsch, J., Silber, J. (Eds.), The Measurement of Individual Well-Being and Group Inequalities. Essays in Memory of Z.M. Berrebi. Routledge, Abingdon, pp. 298–311.
Qizilbash, M., 2004. On the arbitrariness and robustness of multidimensional poverty rankings. J. Hum. Dev. 5, 355–375.
Qizilbash, M., Clark, D.A., 2005. The capability approach and fuzzy poverty measures: an application to the South African context. Soc. Indic. Res. 74, 103–139.
Ramos, X., 2008. Using efficiency analysis to measure individual well-being with an illustration for Catalonia. In: Kakwani, N., Silber, J. (Eds.), Quantitative Approaches to Multidimensional Poverty Measurement. Palgrave-Macmillan, Basingstoke, pp. 155–175.
Ramos, X., Silber, J., 2005. On the application of efficiency analysis to the study of the dimensions of human development. Rev. Income Wealth 51, 285–309.
Ravallion, M., 2011a. On multidimensional indices of poverty. J. Econ. Inequal. 9, 235–248.
Ravallion, M., 2011b. The human development index: a response to Klugman, Rodriguez and Choi. J. Econ. Inequal. 9, 475–478.
Ravallion, M., 2012a. Mashup indices of development. World Bank Res. Obs. 27, 1–32.
Ravallion, M., 2012b. Troubling tradeoffs in the human development index. J. Dev. Econ. 99, 201–209.
Rippin, N., 2010. Poverty Severity in a Multidimensional Framework: The Issue of Inequality Between Dimensions. Georg-August-Universität Göttingen, Courant Research Centre, Poverty, Equity and Growth in Developing and Transition Countries: Statistical Methods and Empirical Analysis, Discussion Papers 47.
Robeyns, I., 2005. Selecting capabilities for quality of life measurement. Soc. Indic. Res. 74, 191–215.
Robeyns, I., 2006. The capability approach in practice. J. Polit. Philos. 14, 351–376.
Roche, J.M., 2013. Monitoring progress in child poverty reduction: methodological insights and illustration to the case study of Bangladesh. Soc. Indic. Res. 112, 363–390.
Roelen, K., Gassmann, F., de Neubourg, C., 2010. Child poverty in Vietnam: providing insights using a country-specific and multidimensional model. Soc. Indicat. Res. 98, 129–145.
Rohde, N., Guest, R., 2013. Multidimensional racial inequality in the United States. Soc. Indic. Res. 114, 591–605.
Rothschild, M., Stiglitz, J.E., 1970. Increasing risk: I. A definition. J. Econ. Theory 2, 225–243.
Ruggeri Laderchi, C., Saith, R., Stewart, F., 2003. Does it matter that we do not agree on the definition of poverty? A comparison of four approaches. Oxf. Dev. Stud. 31, 243–274.
Santos, M.E., 2013. Tracking poverty reduction in Bhutan: income deprivation alongside deprivation in other sources of happiness. Soc. Indic. Res. 112, 259–290.
Savaglio, E., 2006a. Multidimensional inequality with variable population size. Economic Theory 28, 85–94.

Savaglio, E., 2006b. Three approaches to the analysis of multidimensional inequality. In: Farina, F., Savaglio, E. (Eds.), Inequality and Economic Integration. Routledge, London, pp. 264–277.
Schokkaert, E., Van Ootegem, L., 1990. Sen's concept of the living standard applied to the Belgian unemployed. Rech. Econ. Louvain 56, 429–450.
Sen, A.K., 1976. Poverty: an ordinal approach to measurement. Econometrica 44, 219–231.
Sen, A.K., 1977. On weights and measures: informational constraints in social welfare analysis. Econometrica 45, 1539–1572.
Sen, A.K., 1978. Ethical measurement of inequality: some difficulties. In: Krelle, W., Shorrocks, A.F. (Eds.), Personal Income Distribution. North-Holland, Amsterdam, pp. 416–431.
Sen, A.K., 1985. Commodities and Capabilities. North Holland, Amsterdam.
Sen, A.K., 1987. The Standard of Living. With contributions by J. Muellbauer, R. Kanbur, K. Hart and B. Williams, edited by G. Hawthorn. Cambridge University Press, Cambridge.
Sen, A.K., 1992. Inequality Reexamined. Clarendon Press, Oxford.
Sen, A.K., 1993. Capability and well-being. In: Nussbaum, M.C., Sen, A.K. (Eds.), The Quality of Life. Clarendon Press, Oxford, pp. 30–53.
Seth, S., 2013. A class of distribution and association sensitive multidimensional welfare indices. J. Econ. Inequal. 11, 133–162.
Sharma, S., 1996. Applied Multivariate Techniques. Wiley, New York.
Silber, J., 2011. A comment on the MPI index. J. Econ. Inequal. 9, 479–481.
Slesnick, D.T., 1989. Specific Egalitarianism and total welfare inequality: a decompositional analysis. Rev. Econ. Stat. 71, 116–127.
Slesnick, D.T., 1993. Gaining ground: poverty in the postwar United States. J. Polit. Econ. 101, 1–38.
Social Inclusion Division, 2014. What is Poverty? http://www.socialinclusion.ie/poverty.htmlaccessed on 29 June 2014.
Spencer, B.D., Fisher, S., 1992. On comparing distributions of poverty gaps. Sankhyā: Ind. J. Stat. Ser. B 54, 114–126.
Srinivasan, T.N., 1994. Human development: a new paradigm or reinvention of the wheel? Am. Econ. Rev. Pap. Proc. 84, 238–243.
Stiglitz, J.E., Sen, A.K., Fitoussi, J.P., 2009. Report by the Commission on the Measurement of Economic Performance and Social Progress. www.stiglitz-sen-fitoussi.fr.
Streeten, P., 1994. Human development: means and ends. Am. Econ. Rev. Pap. Proc. 84, 232–237.
Sugden, R., 1993. Welfare, resources, and capabilities: a review of inequality reexamined by Amartya Sen. J. Econ. Lit. 31, 1947–1962.
Tchen, A.H., 1980. Inequality for distributions with given marginals. Ann. Probab. 8, 814–827.
The Child Poverty Unit, 2014. Child Poverty Act 2010. http://www.legislation.gov.uk/ukpga/2010/9/contents, accessed on 29 June 2014.
Thorbecke, E., 2007. Multidimensional poverty: conceptual and measurement issues. In: Kakwani, N., Silber, J. (Eds.), The Many Dimensions of Poverty. Palgrave Macmillan, Basingstoke, pp. 3–19.
Thorbecke, E., 2011. A comment on multidimensional poverty indices. J. Econ. Inequal. 9, 485–487.
Tobin, J., 1970. On limiting the domain of inequality. J. Law Econ. 13, 263–277.
Tomlinson, M., Walker, R., Williams, G., 2008. Measuring poverty in Britain as a multi-dimensional concept, 1991 to 2003. J. Soc. Policy 37, 597–620.
Townsend, P., 1970. Measures and explanations of poverty in high income and low income countries: the problems of operationalizing the concepts of development, class and poverty. In: Townsend, P. (Ed.), The Concept of Poverty. Heinemann, London, pp. 1–45.
Townsend, P., 1979. Poverty in the United Kingdom: A Survey of Household Resources and Standards of Living. Penguin, Harmondsworth.
Trani, J.-F., Cannings, T.I., 2013. Child poverty in an emergency and conflict context: a multidimensional profile and an identification of the poorest children in Western Darfur. World Dev. 48, 48–70.
Trani, J.-F., Biggeri, M., Mauro, V., 2013. The multidimensionality of child poverty: evidence from Afghanistan. Soc. Indic. Res. 112, 391–416.
Trannoy, A., 2006. Multidimensional egalitarianism and the dominance approach: a lost paradise? In: Farina, F., Savaglio, E. (Eds.), Inequality and Economic Integration. Routledge, London, pp. 284–302.

Tsui, K.Y., 1995. Multidimensional generalizations of the relative and absolute inequality indices: the Atkinson-Kolm-Sen approach. J. Econ. Theory 67, 251–265.
Tsui, K.Y., 1999. Multidimensional inequality and multidimensional generalized entropy measures: an axiomatic derivation. Soc. Choice Welf. 16, 145–157.
Tsui, K.Y., 2002. Multidimensional poverty indices. Soc. Choice Welf. 19, 69–93.
United Nations Development Programme (UNDP), 1997. Human Development to Eradicate Poverty: Human Development Report 1997. Oxford University Press, New York.
United Nations Development Programme (UNDP), 2005. International Cooperation at a Crossroads. Aid, Trade and Security in an Unequal World: Human Development Report 2005. United Nations Development Programme, New York.
United Nations Development Programme (UNDP), 2010. The Real Wealth of Nations: Pathways to Human Development: Human Development Report 2010. Palgrave Macmillan, Basingstoke.
United Nations Development Programme (UNDP), 2013. The Rise of the South: Human Progress in a Diverse World: Human Development Report 2013. United Nations Development Programme, New York.
van Doorslaer, E., Jones, A.M., 2003. Inequalities in self-reported health: validation of a new approach to measurement. J. Health Econ. 22, 61–87.
van Praag, B.M.S., Ferrer-i-Carbonell, A., 2008. A multidimensional approach to subjective poverty. In: Kakwani, N., Silber, J. (Eds.), Quantitative Approaches to Multidimensional Poverty Measurement. Palgrave-Macmillan, Basingstoke, pp. 135–154.
van Praag, B.M.S., Goedhart, T., Kapteyn, A., 1980. The poverty line—a pilot survey in Europe. Rev. Econ. Stat. 62, 461–465.
van Praag, B.M.S., Frijters, P., Ferrer-i-Carbonell, A., 2003. The anatomy of subjective well-being. J. Econ. Behav. Organ. 51, 29–49.
Vélez, C.E., Robles, M., 2008. Determining the parameters of axiomatically derived multidimensional poverty indices: an application based on reported well-being in Colombia. In: Kakwani, N., Silber, J. (Eds.), Quantitative Approaches to Multidimensional Poverty Measurement. Palgrave-Macmillan, Basingstoke, pp. 210–225.
Wagle, U., 2005. Multidimensional poverty measurement with economic well-being, capability, and social inclusion: a case from Kathmandu, Nepal. J. Hum. Dev. 6, 301–328.
Wagle, U., 2008a. Multidimensional poverty: an alternative measurement approach for the United States? Soc. Sci. Res. 37, 559–580.
Wagle, U., 2008b. Multidimensional Poverty Measurement: Concepts and Application. Springer, New York.
Walzer, M., 1983. Spheres of Justice: A Defense of Pluralism and Equality. Basic Books, New York.
Watts, H., 1968. An economic definition of poverty. In: Moynihan, D.P. (Ed.), On Understanding Poverty. Basic Books, New York, pp. 316–329.
Weymark, J.A., 1981. Generalized Gini inequality indices. Math. Soc. Sci. 1, 409–430.
Weymark, J.A., 2006. The normative approach to the measurement of multidimensional inequality. In: Farina, F., Savaglio, E. (Eds.), Inequality and Economic Integration. Routledge, London, pp. 303–328.
Whelan, C.T., Layte, R., Maître, B., Nolan, B., 2001. Income, deprivation, and economic strain. An analysis of the European community household panel. Eur. Sociol. Rev. 17, 357–372.
Whelan, C.T., Lucchini, M., Pisati, M., Maître, B., 2010. Understanding the socio-economic distribution of multiple deprivation: an application of self-organising maps. Res. Soc. Stratif. Mobil. 28, 325–342.
Whelan, C.T., Nolan, B., Maître, B., 2014. Multidimensional poverty measurement in Europe: an application of the adjusted headcount approach. J. Eur. Soc. Policy 24, 183–197.
Williams, B., 1987. The standard of living: interests and capabilities. In: Sen, A.K. (Ed.), The Standard of Living. With contributions by J. Muellbauer, R. Kanbur, K. Hart and B. Williams, edited by G. Hawthorn. Cambridge University Press, Cambridge, pp. 94–102.
Yaari, M.E., 1987. The dual theory of choice under risk. Econometrica 55, 95–115.
Yaari, M.E., 1988. A controversial proposal concerning inequality measurement. J. Econ. Theor. 44, 381–397.

Yalonetzky, G., 2013. Stochastic dominance with ordinal variables: conditions and a test. Econ. Rev. 32, 126–163.

Yalonetzky, G., 2014. Conditions for the most robust multidimensional poverty comparisons using counting measures and ordinal variables. Soc. Choice Welf. http://dx.doi.org/10.1007/s00355-014-0810-2.

Yu, J., 2013. Multidimensional poverty in China: findings based on the CHNS. Soc. Indic. Res. 112, 315–336.

Yule, G.U., 1900. On the association of attributes in statistics: with illustrations from the material of the childhood society, &c. Phil. Trans. Roy. Soc. A 194, 257–319.

Zheng, B., 1993. An axiomatic characterization of the Watts poverty index. Econ. Lett. 42, 81–86.

Zheng, B., 1997. Aggregate poverty measures. J. Econ. Surv. 11, 123–162.

Zheng, B., 1999. On the power of poverty orderings. Soc. Choice Welfare 16, 349–371.

Zheng, B., 2008. Measuring inequality with ordinal data: a note. In: Bishop, J., Zheng, B. (Eds.), Inequality and Opportunity: Papers from the Second ECINEQ Society Meeting, Research on Economic Inequality, vol. 16. Emerald Group Publishing, Bingley, pp. 177–188.

Zoli, C., Lambert, P.J., 2012. Sequential procedures for poverty gap dominance. Soc. Choice Welf. 39, 649–673.

CHAPTER 4

Equality of Opportunity

John E. Roemer*, **Alain Trannoy**†
*Yale University, New Haven, CT, USA
†Aix-Marseille University, EHESS and CNRS, France

Contents

Abstract

The modern formulation of equality of opportunity emerges from discussions in political philosophy from the second half of the twentieth century beginning with Rawls (1971) and Dworkin (1981a,b). Equality of opportunity exists when policies compensate individuals with disadvantageous circumstances so that outcomes experienced by a population depend only on factors for which persons can be considered

Handbook of Income Distribution, Volume 2A
ISSN 1574-0056, http://dx.doi.org/10.1016/B978-0-444-59428-0.00005-9

to be responsible. Importantly, inequality of opportunity for income exists when individuals' incomes are in some important part determined by the educational achievement and income of the families that raised them. We review the philosophical debates referred to, commenting upon them from an economist's viewpoint. We propose several ways of modeling equality (or inequality) of opportunity, pointing out that an equal-opportunity ethic implies a non-welfarist way of ranking social outcomes. We propose that economic development should be conceived of as the equalization of opportunities for income in a country. We consider equalization of opportunity from a dynamic viewpoint, and we review popular attitudes with regard to distributive justice, showing that there is substantial popular support for an equal-opportunity ethic. We discuss the empirical issues that emerge in measuring inequality of opportunity and provide a review of the empirical literature that measures degrees of inequality of opportunity for the achievement of various objectives, in various countries.

Keywords

Equality of opportunity, Responsibility, Circumstances, Effort, Compensation, Distributive justice

JEL Classification Codes

D63, D3

4.1. INTRODUCTION

In the welfarist tradition of social-choice theory, egalitarianism means equality of welfare or utility.[1] Conservative critics of egalitarianism rightly protest that it is highly questionable that this kind of equality is ethically desirable, as it fails to hold persons responsible for their choices, or for their preferences, or for the way they process outcomes into some interpersonally comparable currency that one can speak of equalizing. In political philosophy, beginning with Rawls (1958, 1971), this critique was taken seriously, and a new approach to egalitarianism transpired, which inserted personal responsibility as an important qualifier of the degree of equality that is ethically desirable. Thus, the development of egalitarian theory, since Rawls, may be characterized as an effort to replace equality of outcomes with equality of opportunities, where opportunities are interpreted in various ways. Metaphors associated with this view are "leveling the playing-field," and "starting gate equality." The main philosophical contributions to the discussion were, following Rawls, from Sen (1980), Dworkin (1981a,b), Arneson (1989), and Cohen (1989).[2] The debate is said to be about "equality of what," and the philosophical view is sometimes called "luck egalitarianism," a term coined by Anderson (1999).

[1] Welfarism is the view that social welfare (or the social objective function) should be predicated only on the utility levels of individuals; that is, that the only information required to compare social alternatives is that summarized in the utility-possibilities sets those alternatives generate. It is a special case of consequentialism. See Chapter 2 for further discussion.

[2] The philosophical literature generated by these pioneers is too large to list here. Book-length treatments that should be mentioned are Rakowski (1993), Van Parijs (1997), and Hurley (2003).

Economists (besides Sen) have been involved in this discussion from 1985 onward. Roemer (1993, 1998) proposed an algorithm for calculating policies that would equalize opportunities for achievement of a given objective in a population. Marc Fleurbaey and François Maniquet contributed economic proposals beginning in the 1990s, and recently summarized in Fleurbaey (2008). Other authors who have contributed to the theory include Van de gaer (1993), Bossert (1995, 1997), and Peragine (2004). An empirical literature is rapidly developing, calculating the extent to which opportunities for the acquisition of various objectives are unequal in various countries, and whether people hold views of justice consonant with equality of opportunity (EOp).

There are various ways of summarizing the significance of these developments for the economics of inequality. Prior to the philosophical contributions that ignited the economic literature that is our focus in this chapter, there was an earlier skirmish around the practical import of equalizing opportunities. Just prior to the publication of Rawls's magnum opus (1971), contributions by Jensen (1969) and Herrnstein (1971) proposed that inequality was in the main due to differential intelligence (IQ), and so generating a more equal income distribution by equalizing opportunities (for instance, through compensatory education of under-privileged children) was a chimera. Economists Bowles (1973) and Conlisk (1974) disagreed; Bowles argued that inequality of income was almost all due to unequal opportunities, not to the heritability of IQ. Despite this important debate on the degree to which economic inequality is immutable, prior to Rawls, economists' discussions of inequality were in the main statistical, focusing on the best ways of measuring inequality.

The post-Rawls–Dworkin inequality literature changed the focus by pointing out that only some *kinds* of inequality are ethically objectionable, and to the extent that economists ignore this distinction, they may be measuring something that is not ethically salient. This distinction between morally acceptable and unacceptable inequality is perhaps the most important contribution of philosophical egalitarian thought of the last 40 years. From the perspective of social-choice theory, equal-opportunity theory has sharply challenged the welfarist assumption that is classically ubiquitous, maintaining that more information than final outcomes in terms of welfare is needed to render social judgment about the ranking of alternative policies—in particular, one must know the extent to which individuals are responsible for the outcomes they enjoy—whether those outcomes were determined by social (and perhaps genetic) factors beyond their control, or not—and this is non-welfare information.

One must mention that another major non-welfarist theory of justice, but a nonegalitarian one, was proposed by Nozick (1974) who argued that justice could not be assessed by knowing only final outcomes; one had to know the process by which these outcomes were produced. His neo-Lockean view, which proposed a theory of the moral legitimacy of private property, can evaluate the justness of final outcomes only by knowing whether the history that produced them was unpolluted by extortion, robbery, slavery, and so on.

Simply knowing the distribution of final outcomes (in terms of income, welfare, or whatever) does not suffice to pass judgment on the distribution's moral pedigree. So the period since 1970 has been one in which, in political philosophy, non-welfarist theories flourished, on both the right and left ends of the political spectrum.

In this chapter, we begin by summarizing the philosophical debate concerning equality since Rawls (Section 4.2), presenting economic algorithms for computing policies which equalize opportunities—or, more generally, ways of ordering social policies with respect to their efficacy in opportunity equalization (Sections 4.3, 4.4, and 4.5), application of the approach to the conceptualization of economic development (Section 4.6), discussion of dynamic issues (Section 4.7), a preamble to a discussion of empirical work (Section 4.8), evidence of population views from surveys and experiments concerning conceptions of equality (Section 4.9), a discussion of measurement issues, and summary of the empirical literature on inequality of opportunity to date (Section 4.11). We conclude by mentioning some critiques of the equal-opportunity approach, and some predictions (Section 4.12).

4.2. EGALITARIAN POLITICAL PHILOSOPHY SINCE RAWLS

Rawls (1958) first published his ideas about equality over 50 years ago, although his magnum opus did not appear until 1971. His goal was to unseat utilitarianism as the ruling theory of distributive justice, and to replace it with a type of egalitarianism. He argued that justice requires, after guaranteeing a system which maximizes civil liberties, a set of institutions that maximize the level of "primary goods" allocated to those who are among the worse off in society, in the sense of receiving the least amount of these goods. Economists call this principle maximin primary goods; Rawls often called it the difference principle. Moreover, he attempted to provide an argument for the recommendation based on construction of a "veil of ignorance" or "original position," which shielded decision makers from knowledge of information about their situations that was "morally arbitrary," so that the decision they came to regarding just allocation would be impartial. Thus, Rawls's (1971) project was to derive principles of justice from rationality and impartiality.

Rawls did not advocate maximining utility (even assuming interpersonal utility comparisons were available), but rather maximining (some index of) primary goods. This was, in part, his attempt to embed personal responsibility into the theory. For Rawls, welfare was best measured as the extent to which a person is fulfilling his or her plan of life: but he viewed the choice of life plan as something up to the individual, upon which social institutions had no business passing judgment. Primary goods were deemed to be those inputs that were required for the success of any life plan, and so equalizing primary goods bundles across persons (or passing to a maximin allocation that would dominate component-wise an equal allocation) was a way of holding persons responsible for their life-plan choice. The question of how to aggregate the various primary goods into an index that would allow comparison of bundles was never successfully solved by Rawls (and some skeptical economists said that the subjective utility function was the obvious way to aggregate primary goods).

Rawls defended the difference principle by arguing that it would be chosen by decision makers who were rational but were deprived of knowledge about their own situations in the world, to the extent that this knowledge included information about their physical, social, and biological endowments, which were a matter of luck, and therefore whose distribution Rawls described as morally arbitrary. He named the venue in which these souls would cogitate about justice the "original position." In the original position, souls were assumed to know the laws of economics and to be self-interested. They were, moreover, to be concerned with the allocation of primary goods, because they did not know their life plans, or even the *distribution* of life plans in the actual society. Nor were they to know the *distribution* of physical and biological endowments in society.

Here, we believe Rawls made a major conceptual error. If the veil of ignorance is intended to shield decision makers from knowledge of aspects of their situations that are morally arbitrary, and only of those aspects, they *should* know their plans of life, which, by hypothesis, are not morally arbitrary, because Rawls deems that persons are responsible for their life plans. Second, although a person's *particular* endowment of resources, natural and physical, might well be morally arbitrary (to the extent that these were determined by the luck of the birth lottery), the *distribution* of these resources is a fact of nature and society, and should be known by the denizens in the original position, just as they are assumed to know the laws of economics. Therefore, Rawls constructed his veil too thickly, on two counts, given his philosophical views.

Given the paucity of information available to the decision makers in the original position, it is not possible to use classical decision theory to solve the problem of the desirable allocation of primary goods. Indeed, the only precise arguments that Rawls gives for the conclusion that the difference principle would be chosen in the original position occur at Rawls, 1999[1971], p. 134), and they essentially state that decision makers are extremely risk averse. For example:

> *The second feature that suggests the maximin rule is the following: the person choosing has a conception of the good such that he cares very little, if anything, for what he might gain about the minimum stipend that he can, in fact, be sure of by following the maximin rule. It is not worthwhile for him to take a chance for the sake of further advantage, especially when it may turn out that he loses much that is important to him. The last provision brings in the third feature, namely, that the rejected alternatives have outcomes one can hardly accept. The situation involves grave risks.*

But extreme risk aversion, which Rawls here depends on for his justification of maximin, is certainly not an aspect of rationality.

Thus, despite its enormous influence in political philosophy, Rawls's argument for maximin is marred in two ways: first, its reliance on deducing the principle of justice from the original position was crucially flawed in depriving the denizens of that position of knowledge of features of themselves (life plans) and of the world (the distributions of various kinds of resources, including genetic ones, and ones possessed by families into which a person is born) which were *not* morally arbitrary,[3] and second, for its assumption (despite claims to the contrary by Rawls and others) that decision makers were extremely risk averse. The value of Rawls's contribution is in stating a radical egalitarian position about the injustice of receiving resources through luck—and, in particular, the luck of the birth lottery—and that it shifted the equalisandum from utility to a kind of resource, primary goods. In our view, however, the project of deducing equality or maximin from rationality and impartiality alone was a failure. Indeed, Moreno-Ternero and Roemer (2008) argue that some solidaristic postulate is necessary to deduce maximin or, more generally, to deduce some kind of egalitarianism as the ordering principle for social choice. Although egalitarians might wish to deduce their view from postulates that can garner universal approval (like rationality and impartiality), this is not possible. Therefore, an egalitarian theory of justice cannot have *universal* appeal, if the solidaristic postulate, which we believe necessary, is contentious.

Although Rawls is usually viewed as the most important egalitarian political philosopher of the twentieth century, one may challenge the claim that his view is egalitarian: to wit, the just income distribution, for Rawls, allows incentive payments to the highly skilled in order to elicit their productive activity, even though this produces inequality. The main philosopher who challenges Rawls's acceptance of incentive-based income inequality is Cohen, whom we discuss below.

In 1981, Ronald Dworkin published two articles that essentially addressed the problems in the Rawlsian argument that we have summarized, although he did not use the Rawlsian language (original position, primary goods). His project was to define a conception of equality that was ethically sound. In the first of these articles, he argued that "equality of welfare" was not a sound view, primarily because equality of welfare does not hold persons responsible for their preferences. In particular, Dworkin argued that if a person has expensive tastes, and he identifies with those tastes, society does not owe him an additional complement of resources to satisfy them. (The only case of expensive tastes, says Dworkin, that justifies additional resources are those tastes that are addictions or compulsions, tastes with which the person does not "identify," and would prefer he

[3] We reiterate it is the distribution of traits that is a fact of nature, and hence not morally arbitrary, whereas the endowment of a given individual may well be morally arbitrary, in the sense of its being due to luck.

did not have.) In the second article, Dworkin argues for "equality of resources," where resources include (as for Rawls) aspects of a person's physical and biological environment for which he should not be held responsible (such as those acquired through birth).

But how can one "equalize resources," when these comprise both transferable goods, such as money, and inalienable resources, including talent, families into which persons are born, and even genes? Dworkin proposed an ingenious device, an insurance market carried out behind a veil of ignorance, where the "souls" participating represent actual persons and know the preferences of those whom they represent, but do not know the resources with which their persons are actually endowed in the world. In this insurance market, each participant would hold an equal amount of some currency and would be able to purchase insurance with that currency against bad luck in the birth lottery, that is, the lottery in which nature assigns souls to persons in the world (or resource endowments to souls). Dworkin argues that the allocation of goods that would be implemented after the birth lottery occurred, the state of the world was revealed, and insurance policies taken behind the Dworkinian veil were settled, was an allocation that "equalized resources." It held persons responsible for their preferences—in particular, their risk preferences—and was egalitarian because all souls were endowed, behind the veil, with the same allotment of currency with which to purchase insurance. Impartiality with respect to the morally arbitrary distribution of resources was accomplished by shielding the souls from knowledge of their endowments in the actual world associated with the birth lottery (genetic and physical). Thus, Dworkin retained Rawls's radical egalitarian view about the moral arbitrariness of the distribution of talents, handicaps, and inherited wealth, but implemented a mechanism that held persons responsible for their tastes that was much cleaner than discarding preferences and relying on primary goods, as Rawls had done.

Despite the cleverness of Dworkin's construction, it can lead to results that many egalitarians would consider perverse. To illustrate the problem, consider the following example. Suppose there are two individuals in the world, Andrea and Bob. Andrea is lucky: she has a fine constitution and can transform resources (wealth) into welfare at a high rate. Bob is handicapped; his constitution transforms wealth into welfare at exactly one-half of Andrea's rate. We assume, in particular, that Andrea and Bob have interpersonally comparable welfare. The internal resource that Andrea possesses and Bob lacks is a fine biological constitution (say, a healthy supply of endorphins).

We assume that Bob and Andrea have the same risk preferences over wealth: They are each risk averse and have the von Neumann–Morgenstern utility function over wealth $u(W) = \sqrt{W}$. Suppose that the distribution of (material) wealth in the world to (Andrea, Bob) would be (W^A, W^B), with no further intervention. Thus each individual is endowed with an internal constitution and some external resource.

We construct Dworkin's hypothetical insurance market as follows.[4] Behind the veil of ignorance, there is a soul, Alpha, who represents Andrea, and a soul, Beta, who represents Bob. These souls know the risk preferences of their principals and the constitutions of Andrea and Bob, but they do not know which person they will become in the birth lottery. Thus, from their viewpoints, there are two possible states of the world, summarized in the table:

State 1	Alpha becomes Andrea	Beta becomes Bob
State 2	Alpha becomes Bob	Beta becomes Andrea

Each state occurs with probability one-half. *We* know that state 1 will indeed occur, but the souls face a birth lottery with even chances, in which they can take out insurance against bad luck (that is, of becoming Bob).

There are two commodities in the insurance market: a commodity x_1, a unit of which pays the owner \$1 if state 1 occurs, and a commodity x_2, a unit of which pays \$1 if state 2 occurs. Each soul can either purchase or sell these commodities: selling one unit of the first commodity entails a promise to deliver \$1 if state 1 occurs. Each soul possesses, initially, zero income (behind the veil) with which to purchase these commodities. In particular, they have *equal wealth endowments* behind the veil in the currency that is recognized in that venue. Thus, the insurance market acts to redistribute tangible wealth in the actual world to compensate persons for their natural endowments, which cannot be altered, in that way which the souls, who represent persons, would desire, had they been able to insure against the luck of the birth lottery. It is an institution that transforms what Dworkin calls "brute luck" into "option luck." The former is luck that is not insurable; the latter is luck whose outcome is protected by insurance, or the outcome of a gamble one has chosen to take.

An equilibrium in this insurance market consists of prices $(1,p)$ for commodities (x_1,x_2), demands $(x_1^\alpha,x_2^\alpha),(x_1^\beta,x_2^\beta)$ by souls Alpha and Beta for the two contingent commodities, such that

(1) $\left(x_1^\alpha, x_2^\alpha\right)$ maximizes $\frac{1}{2}\sqrt{W^{\mathrm{A}}+x_1^\alpha}+\frac{1}{2}\sqrt{\frac{W^{\mathrm{B}}+x_2^\alpha}{2}}$
subj. to $x_1^\alpha + px_2^\alpha = 0$

(2) $\left(x_1^\beta, x_2^\beta\right)$ maximizes $\frac{1}{2}\sqrt{W^{\mathrm{B}}+x_1^\beta}+\frac{1}{2}\sqrt{2\left(W^{\mathrm{A}}+x_2^\beta\right)}$
subj. to $x_1^\beta + px_2^\beta = 0$

(3) $x_s^\alpha + x_s^\beta = 0 \quad \text{for } s=1,2$

[4] Dworkin did not propose a formal model but relied on intuition. The model here is a version of an Arrovian market for contingent claims.

Let us explain these conditions. Condition (1) says that Alpha chooses her demand for contingent commodities optimally, subject to her budget constraint—that is, she maximizes her expected utility. Her utility if she becomes Andrea (state 1), will be $\sqrt{W_1^{\mathrm{A}} + x_1^{\alpha}}$. Now if Alpha becomes Bob (state 2), her wealth will be $W^{\mathrm{B}} + x_2^{\alpha}$; however, from the viewpoint of her principal, Andrea, that will generate only half as much welfare, so she evaluates this wealth as being worth, in utility terms, $\sqrt{\frac{W^{\mathrm{B}} + x_2^{\alpha}}{2}}$. Condition (2) has a similar derivation, but this time, soul Beta takes the benchmark situation as becoming Bob. Condition (3) says that both markets clear.

The equilibrium is given by

$$p = 1, \quad \left(x_1^{\alpha}, x_2^{\alpha}\right) = \left(\frac{2W^{\mathrm{B}} - W^{\mathrm{A}}}{3}, \frac{W^{\mathrm{A}} - 2W^{\mathrm{B}}}{3}\right),$$

$$\left(x_1^{\beta}, x_2^{\beta}\right) = \left(\frac{-2W^{\mathrm{B}} + W^{\mathrm{A}}}{3}, \frac{-W^{\mathrm{A}} + 2W^{\mathrm{B}}}{3}\right).$$

Now state 1 occurs. Therefore Andrea, after the insurance contracts are settled, ends up with wealth $W^{\mathrm{A}} + x_1^{\alpha} = \frac{2}{3}(W^{\mathrm{A}} + W^{\mathrm{B}})$—two-thirds of the total wealth—and Bob ends up with one-third of the total wealth. The result is perverse because *Bob is the one with the low resource endowment*, that is, with a low ability to transform money into welfare. It is Bob, putatively, whom an equal-resource principle should compensate, but it is Andrea who ends up the winner.[5] Even should state 2 have occurred, the outcome would have been the same—two-thirds of the wealth would end up being Andrea's.

Why does this happen? Because, even though both souls are risk averse, they are not sufficiently risk averse to induce them to shift wealth into the bad state (of being born Bob); it is more worthwhile (in terms of expected utility) to use wealth in the state when it can produce a lot of welfare (when a soul turns out to be Andrea). If the agents were *sufficiently* risk averse, this would not occur. (If the utility function were $u(W) = W^c/c$, and $c < 0$, then, post-insurance, Bob would end up with more wealth than Andrea. If the utility function is $u(W) = \log W$, then the agents split the wealth equally.) But the example shows that in general the hypothetical insurance market does not implement the kind of compensation that Dworkin desires: for Bob is the one who suffers from a deficit in an internal resource—from morally arbitrary bad luck. For Dworkin's insurance market to avoid this kind of perversity, individuals would have to be sufficiently risk averse, and this

[5] This perversity of the Dworkin insurance mechanism was first pointed out by Roemer (1985). Dworkin never proposed a model of the insurance market but conjectured that it would reallocate wealth in a way to compensate those with a paucity of nontransferable resources. He continued to use the insurance-market thought experiment to justify social policies (e.g., in the case of national health insurance for the United States), even though his thought experiment did not necessarily produce the compensatory redistributions that he thought it would implement.

it is inappropriate to assume, for the theory should surely produce the desired result (of compensating those with a paucity of internal resources) in the special case that all agents have the same risk preferences.[6]

In the model just presented of the hypothetical insurance market, note that it was necessary to make interpersonal welfare comparisons. Alpha, Andrea's soul, has to contemplate how she would feel if she were to be born as Bob, and with a given amount of wealth. She does this by transforming Bob's wealth into a *welfare-equivalent wealth* for Andrea. And soul Beta has to make a similar interpersonal comparison. We maintain that it is impossible to construct a veil-of-ignorance thought experiment without making such comparisons. The point is simple: if a soul has to compare how it would feel when being incarnated as different persons, it must be able to make interpersonal welfare comparisons. Without the ability to compare the lives of different persons in different circumstances, an investment in insurance would have no basis.[7]

Despite the problem we have exhibited with Dworkin's proposal, it was revolutionary, in the words of Cohen, in transporting into egalitarian theory the most powerful tool of the anti-egalitarian Right, the importance of personal responsibility. One might argue, after seeing the above demonstration, that Dworkin's insurance market is an appealing thought experiment, and therefore one should give up on the egalitarian impulse of compensating persons for features of their situations for which they are not responsible; that is, instead of rejecting Dworkin's model as inadequate, one should reject his egalitarian desideratum. Moreno-Ternero and Roemer (2008) consider this and argue instead that the veil of ignorance is an inappropriate thought experiment for ascertaining what justice requires. Although their arguments for this are new, the position is not: It was also advocated earlier by Barry (1991).

In the example we have given, there is, for egalitarians, a moral requirement to transfer tangible wealth from Andrea to Bob, because Bob lacks an inalienable resource that Andrea possesses, the ability effectively to transform goods into welfare, a lack that is beyond his control, and due entirely to luck. Dworkin also focused upon a different possible cause of unequal welfares, that some persons have expensive tastes, while others have cheap ones. His view was that persons with expensive tastes do *not* merit additional

[6] When Dworkin was confronted with this example at a conference in Halifax in 1985, he responded that he would not use the insurance device in cases where it produced the "pathological" result. This is, however, probably an unworkable position, for how does one characterize a priori the set of admissible economic environments?

This is not the first time that insufficient concavity of preferences causes problems for economic analysis. See, for example, the discussion of money-metric utility in Chapter 1.

[7] Readers may recall that Harsanyi (1955) claimed to construct a veil-of-ignorance argument for utilitarianism without making interpersonal comparisons. But his argument fails—not as a formal mathematical statement, but in the claim that utilitarianism is what has been justified. (See, for an early discussion, Weymark, 1991; and for a more recent one, Moreno-Ternero and Roemer, 2008.)

wealth in order to satisfy them, as long as those persons were satisfied with their tastes, or, as he said, identified with them. There is no injustice in a world where wealth is equal, but those with champagne tastes suffer compared to those with beer tastes, due to the relative consumptions of champagne and beer that that equal wealth permits. So the "pathology" that we have illustrated with the Andrea and Bob example depends on the source of Bob's relative inefficiency in converting wealth into welfare being a handicap, rather than an expensive taste.

Slightly before Dworkin's articles were published, Sen (1980) gave a lecture in which he argued that Rawls's focus on primary goods was misplaced. Sen argued that Rawls was "fetishist" in focusing on goods, and should instead have focused on what the goods provide for people, which he called "functionings"—being able to move about, to become employed, to be healthy, and so on. Sen defined a person's *capability* as the set of vectors of functionings that were available to him, and he called for equality of capabilities.[8] Thus, although a rich man on a hunger strike might have the same (low) functioning as a poor man starving, their capabilities are very different. While not going so far as to say utilities should be equalized, Sen defined a new concept between goods and welfare—functionings—which Cohan (1993) later described as providing a state of being that he called "midfare." For Sen, the opportunity component of the theory was expressed in an evaluation not of a person's actual functioning level, but of what functionings were *available* to him, his "capability."

Sen's contribution led to both theoretical and practical developments. On the theoretical level, it inspired a literature on comparing opportunity (or feasible) sets: If one desires to "equalize" capabilities, it helps to have an ordering on sets of sets. See Foster's (2011) summary of this literature. On the practical side, it led to the human development index, published annually by the UNDP. For development of Sen's capability approach, see Chapter 2.

Later in the decade, further reactions to Dworkin came from philosophers, notably Arneson (1989) and Cohen (1989). Arneson argued that Dworkin's expensive-taste argument against equality-of-welfare was correct, but his alternative of seeking equality of resources was not the only option: Instead, one should seek to equalize *opportunities for welfare*. This, he argued, would take care of the expensive-tastes problem. Rather than relying on the insurance mechanism to define what resource egalitarianism means, Arneson proposed to distribute resources so that all persons had equal opportunity for welfare achievement, although actual welfares achieved would differ because people would make different choices. There are problems with formalizing Arneson's proposal (see Roemer, 1996), but it is notable for not relying on any kind of veil of ignorance, in contrast to the proposals of Rawls and Dworkin.

[8] Sen has not proposed an ordering of sets that would enable one to compare capabilities.

Cohen (1989) criticized Dworkin for making the wrong "cut" between resources and preferences. The issue, he said, was what people should or should not be held responsible for. Clearly, a person should not be held responsible for his innate talents and inherited resources, but it is not true that a person should be fully responsible for his preferences either, because preferences are to some (perhaps large) degree formed in circumstances (in particular, those of one's childhood), which are massively influenced by resource availability. Indeed, if a person has an expensive taste for champagne due to a genetic abnormality, he would merit compensation under an egalitarian ethic.[9] Cohen's view was that inequality is justified if and only if it is attributable to choices that are ones for which persons can sensibly he held responsible—so if a person who grows up poor, develops a "taste" against education, induced by the difficulty of succeeding in school due to lack of adequate resources—a taste with which he even comes to "identify"—then Cohen would not hold him responsible for the low income due to his consequently low wage, while Dworkin presumably would hold him responsible. Cohen does not propose a mechanism or algorithm for finding the just distribution of resources, but provides a number of revealing examples (see, for example, Cohen, 1989, 2004). He calls his approach "equal access to advantage."

Besides criticizing Dworkin for his partition of the space of attributes and actions into ones for which compensation is, or is not, due, Cohen (1997), importantly, critiqued Rawls's difference principle, as insufficiently egalitarian. The argument is based on Rawls's restriction of the ambit of justice to the design of social institutions—in particular, that ambit does not include personal behavior. Thus, the Rawlsian tax system should attempt to maximize the welfare of the least-well-off group in society, under the assumption that individuals choose their labor supplies to maximize their personal utility. Suppose the highly skilled claim that if their taxes are raised from 30% to 50%, they will reduce their labor supply so much that the worst-off group would be less well off than it is at the 30% tax rate. If 30% is the tax rate that maximizes the welfare (or income) of the least well off, given this self-interested behavior of the highly skilled, then it is the Rawlsian-just rate. But Cohen responds that, as long as the highly skilled are at least as well off as the worst off at the 50% tax rate, then justice requires the 50% tax rate. This difference of viewpoint between Rawls and Cohen occurs because Cohen requires individuals to act, in their personal choices, according to the commands of the difference principle (that is, to take those actions that render those who are worst off to be as well off as possible), and Rawls does not. Indeed, Rawls stipulates that one requirement of a just society is that its members endorse the conception of justice. It is peculiar, Cohen remarks, that that conception should apply only to the design of social institutions, and not to personal behavior.

[9] This is not a crazy example. There is a medically recognized syndrome in which people who sustain a certain kind of brain injury come to crave expensive foods (see Otsuka (2011, p. 81).

A question that arises from the discussion of responsibility is its relationship to freedom of the will. If responsibility has become central in the conceptualization of just equality, does one have to solve the problem of free will before enunciating a theory of distributive justice? Different answers are on offer. We believe the most practical answer, which should suffice for practicing economists, is to view the degree of responsibility of persons as a parameter in a theory of equality. Once we assign a value to this parameter, then we have a particular theory of EOp, because we then know for what to hold persons responsible. The missing parameter is supplied by each society, which has a concept of what its citizens should be held responsible for; hence, there is a specific theory of EOp for each society, that is, a theory that will deliver policy recommendations consonant with the theory of responsibility that that society endorses. This is a political approach, rather than a metaphysical one.

Another answer to the free will challenge is to make a distinction prevalent among philosophers. "Compatibilists" are those philosophers who believe that it is consistent both to endorse determinism (in the sense of a belief in the physical causation of all behavior) *and* the possibility of responsibility; incompatibilists are those who believe that determinism precludes responsibility. Most philosophers (who think about the problem) are probably, at present, compatibilists. For instance, Scanlon (1986) believes that the determinist causal view is true, but also that persons can be held responsible for their behavior, as long as they have contemplated their actions, weighed alternatives, and so on. (The issue of sufficient contemplation is independent of the issue of the cause of expensive tastes, raised above.) From a practical viewpoint, the problem of free will therefore does not pose a problem for designing policies motivated by the idea that persons should not be held accountable for aspects of their condition that are due to circumstances beyond their control.

The philosophical literature on "responsibility-sensitive egalitarianism" continues beyond the point of this quick review, but enough summary has been provided to proceed to a discussion of economic models.

4.3. A MODEL AND ALGORITHM FOR EQUAL-OPPORTUNITY POLICY

Consider a population whose members are partitioned into a finite set of *types*. A type comprises the set of individuals with the same circumstances, where *circumstances* are those aspects of one's environment (including, perhaps, one's biological characteristics) which are beyond one's control, and influence outcomes of interest. Denote the types $t = 1, \ldots. T$. Let the population fraction of type t in the population be f^t. There is an *objective* for which a planner wishes to equalize opportunities. The degree to which an individual will achieve the objective is a function of her circumstances, her *effort*, and the social policy: We write the value of the objective as $u^t(e, \varphi)$, where e is a measure of effort and $\varphi \in \Phi$ the set of social policies. Indeed, $u^t(e, \varphi)$ should be considered to be the average achievement of the objective among those of type t expending effort e when the policy is φ. Here, we will take effort to be a nonnegative real number. Later, we will introduce luck into the problem.

u^t is not, in general, a subjective utility function: Indeed u^t is assumed to be monotone *increasing* in effort, whereas subjective utility is commonly assumed to be decreasing in standard conceptions of effort. Thus, u might be the adult wage, circumstances could include several aspects of childhood and family environment, and e could be years of schooling. Effort is assumed to be a choice variable for the individual, although that choice may be severely constrained by circumstances, a point to which we will attend below. The final data for the problem consist of the distributions of effort within types as a function of policy: For the policy φ, denote the distribution function of effort in type t as $G_\varphi^t(\cdot)$. We would normally say that effort is chosen by the individual by maximizing a preference order, but preferences are not the fundamentals of this theory: Rather, the data are $\{T, G_\varphi^t, f^t, u, \Phi\}$, where we use T to denote, also, the set of types.

Defining the set of types and the conception of effort assumes that the society in question has a conception of the partition between responsible actions and circumstances, with respect to which it wishes to compute a consonant approach to equalizing opportunities. We describe the approach of Roemer (1993, 1998). The verbal statement of the goal is to find that policy which nullifies, to the greatest extent possible, the effect of circumstances on outcomes, but allows outcomes to be sensitive to effort. Effort comprises those choices that are thought to be the person's responsibility. But note that the *distribution* of effort in a type at a policy, G_φ^t, is not due to the actions of any person (assume here a continuum of agents), but is a characteristic of the type. If we are to indemnify individuals against their circumstances, we must not hold them responsible for being members of a type with a poor distribution of effort.

We require a measure of *accountable* effort, which, because effort is influenced by circumstances, cannot be the raw effort e. (Think of years of education—raw effort—which is surely influenced in a major way by social circumstances.) Roemer proposed to measure accountable effort as the rank of an individual on the effort distribution of her type: thus, if for an individual expending effort e, $G_\varphi^t(e) = \pi$, we say the individual expended the *degree* of effort π, as opposed to the *level* of effort e. The rank provides a way of making inter-type comparisons of the efforts expended by individuals. A person is judged accountable, that is to say, by comparing her behavior only to others with her circumstances. In comparing the degrees of effort of individuals across types, we use the rank measure, which sterilizes the distribution of raw effort of the influence of circumstances upon it.[10]

[10] Some authors (Ramos and Van de gaer, 2012) have called this move—of identifying the degree of effort with the rank of the individual on the objective distribution of his type—the Roemer identification axiom (RIA). Although the name is lofty, the idea is simple: Persons should not be held responsible for characteristics of the distribution of effort in their type, for that distribution is a circumstance.

Because the functions u^t are assumed to be strictly monotone increasing in e, it follows that an individual will have the same rank on the distribution of the objective, within his type, as he does within the distribution of effort of his type.[11] Define:

$$v^t(\pi,\varphi) = u^t(e^t(\pi),\varphi)$$

where $e^t(\pi)$ is the level of effort at the πth quantile of the distribution G^t_φ, that is, $G^t_\varphi(e^t(\pi)) := \pi$. Then the functions $v^t(\cdot,\varphi)$ are the inverse functions of the distribution functions of the objective, by type, under the policy φ. (In this sense, v^t is like Pen's parade, which is also the inverse of a distribution function.) Inequality of opportunity holds when these *functions* are not identical. In particular, because we are viewing persons at a given rank π as being equally accountable with respect to the choice of effort, the vertical difference between the functions $\{v^t(\cdot,\varphi)\}$ is a measure of the extent of inequality of opportunity (or, equivalently, the horizontal distance between the cumulative distribution functions).

What policy is the optimal one, given this conception? We do not simply want to render the functions v^t identical at a low level, so we need to adopt some conception of "maxi-minning" these functions. We want to choose that policy which pushes up the lowest v^t function as much as possible—and as in Rawlsian maximin, the "lowest" function may itself be a function of what the policy is. A natural approach is therefore to maximize the area below the lowest function v^t, or more precisely, to find that policy which maximizes the area under the *lower envelope* of the functions $\{v^t\}$. The formal statement is to:

$$\max_{\varphi\in\Phi}\int_0^1 \min_t v^t(\pi,\varphi)\mathrm{d}\pi. \tag{4.1}$$

We call the solution to this program the opportunity-equalizing policy, φ^{EOp}. (Computing (4.1) is equivalent to maximizing the area to the left of the left-hand envelope of the type distributions of the objective, and bounded above by the horizontal line of height one.)

In the case in which the lower envelope of the functions $\{v^t\}$ is the function of a single type (the unambiguously most disadvantaged type), what we have done is simply to maximize the average value of the objective for the most disadvantaged type, since $\int_0^1 v^t(\pi,\varphi)\mathrm{d}\pi$ is simply the mean value of the objective for type t at policy φ.

Thus, the approach implements the view that differences between individuals caused by their circumstances are ethically unacceptable, but differences due to differential effort are all right. Full EOp is achieved not when the value of the objective is equal for all, but when members of each type face the *same chances*, as measured by the distribution functions of the objective that they face.

[11] If actual effort is a vector, then a unidimensional measure e would be constructed, for example, by regressing the objective values against the dimensions, thus computing weights on the dimensions of raw effort.

One virtue of the approach taken here is that it is easy to illustrate graphically. In Figure 4.1, we present two graphs, to illustrate inequality of opportunity in Hungary and Denmark. In each graph, there are three cumulative income distributions, corresponding to male workers of three types: those whose more educated parent had no more than lower secondary education, those whose more educated parent just completed

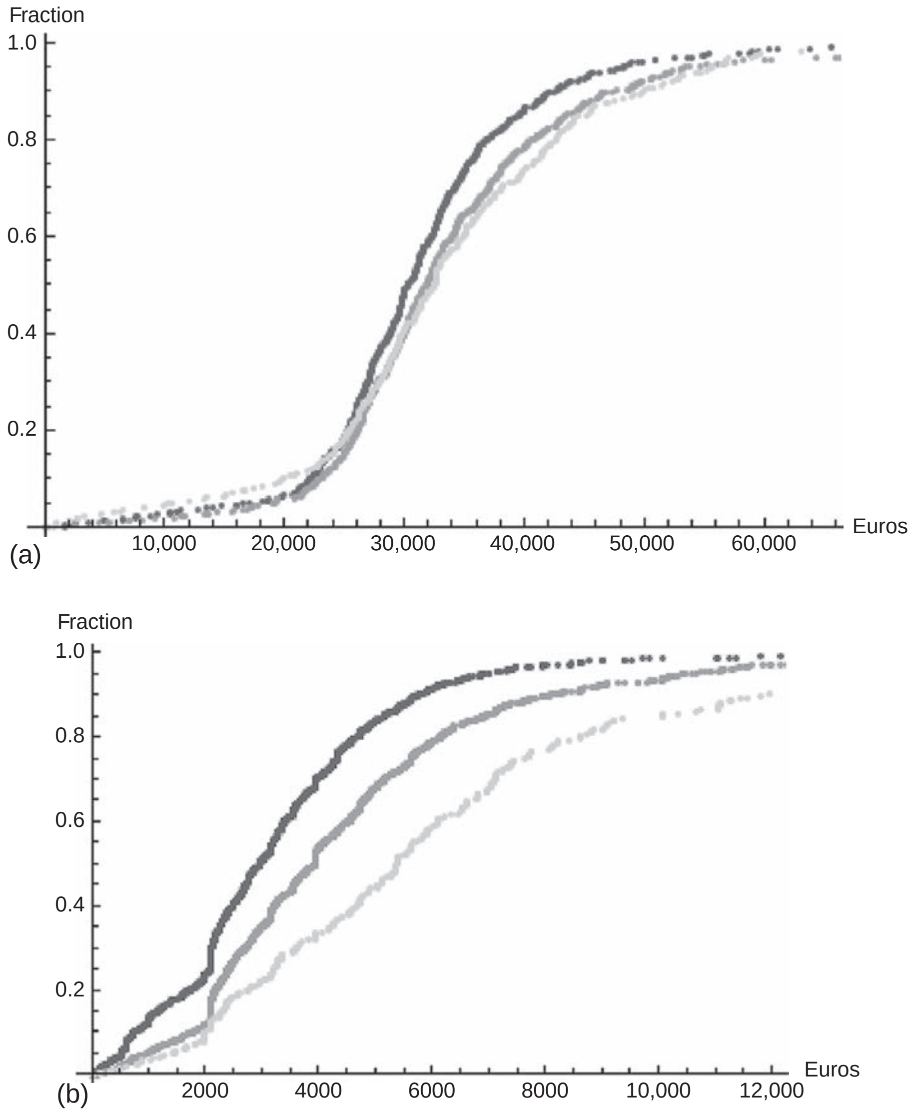


Figure 4.1 (a) Three income distribution functions for Danish male workers, according the circumstance of parental education. (The curve with darkest hue is the income distribution of sons of the least-educated parents; the lightest hue is the income distribution of sons of the most-educated parents.) (b) As in (a), but for Hungary.

secondary education, and those whose more educated parent had at least some tertiary education. (The data are from EU-SILC-2005.) The inverses of these distribution functions are the functions $v^t(\cdot, \varphi)$ defined above. The policy is the status-quo policy. It seems clear that, with respect to this one circumstance (parental education), opportunities for income have been more effectively equalized in Denmark than in Hungary.[12] The graphs are taken from Roemer (2013).

The approach inherent in (4.1) is one that treats all causes of inequality not accounted for by a person's type as being due to effort. For example, with respect to Figure 4.1, there are many circumstances that influence outcomes not accounted for in the definition of type, and so the inequality of opportunity illustrated in that figure should be considered to be a lower bound on the true inequality of opportunity. Nevertheless, it is often the case that delineating only a few circumstances will suffice to illustrate obvious inequality of opportunity, and one can say that social policy should attempt to mitigate at least that inequality.

Let us note that the equal-opportunity approach is *non-welfarist* or more precisely *non-consequentialist*. A welfarist procedure for ordering social policies uses information only in the objective possibilities sets of the population associated with those procedures. In the income example, it would use only the data of the income distribution of the population and ignore the data of what individuals were of what types. Circumstances are non-welfare (or nonobjective) information. More informally, consequentialism only considers the final results of policies (incomes), and not the causes of those consequences. Here, we say there are two kinds of cause of outcomes with different moral status: circumstances and effort. We must distinguish between these causes and social policy should attempt to mitigate the inequality effects of one of them, but not necessarily of the other.

At this point, we return briefly to consider a philosophical critique of this approach—and indeed of the general evolution of responsibility-sensitive egalitarianism, as it was reviewed in Section 4.1 above—offered by Hurley (2002), who writes that "Roemer's account does not show how the aim to neutralize luck could provide a basis for egalitarianism." Hurley says that, absent luck, many possible distributions of the objective could have occurred, and one cannot claim that "neutralizing" luck means to render outcomes sensitive only to degrees of effort. Moreover, she writes that it is not an *argument* for EOp that it neutralizes the effects of luck.

The moral premise of the EOp view is that rewards should be sensitive only to the autonomous efforts of individuals. This is a special case of rewards according to deserts. People deserve, in the EOp view, to acquire the objective in proportion to how hard they try. Thus, strictly speaking, the EOp view is not one whose fundamental primitive is equality: deservingness is fundamental, together with the normative thesis that justified inequality tracks deservingness. Inequalities that are not due to unequal efforts are *defined*

[12] We say "seems" clear, because the horizontal-axis euro scale is different in the two figures.

as being due to luck; that is, luck is so-called because it is a cause of reward that is illegitimate from the EOp view. The statement that "EOp intends to neutralize the effects of luck on outcomes" is therefore equivalent to the statement "EOp intends to render outcomes sensitive only to effort."

So, for example, suppose a child, A, does well in life because his parents were rich, not because he exerted great effort, while another child, B, from a poor family, does well by virtue of exerting great effort. Some might argue that it may be no less a matter of luck that B was the kind of person who works hard than that A had rich parents, but that approach, whatever its merits, is not the sense in which responsibility-concerned egalitarians use the word "luck." Luck, for us, *means* the source of noneffort-caused advantage. To be sure, it is not an *argument* for EOp that it neutralizes luck; it is rather *definitional* of the EOp view that it does so. The *argument* for EOp must be that it is *right* to render outcomes sensitive only to effort.[13]

The next example, which is hypothetical, is given to illustrate the difference between the equal-opportunity approach and the approach that is conventional in many areas of social policy, utilitarianism. A *utilitarian* policy maximizes the average value of the objective in a population. Utilitarianism is a special case of welfarism, although there are many welfarist preference orderings of policies.

We consider a population partitioned into T types, where the frequency of type t is f^t. The population suffers from I diseases, with the generic disease denoted i. The types might be defined by socioeconomic characteristics,[14] and the Health Ministry is interested in mitigating the effect of socioeconomic characteristics on health. There is available in the health sector an amount of resource (money), $\overline{R}$ per capita. We do not address how much of a society's product should be dedicated to health, but only how to spend the amount that has been so dedicated. Effort is here conceived of as lifestyle quality (exercise, smoking behavior, etc.). We choose the policy space to be allocations of the resource to treating various diseases; that is vectors $R=(R^1, \ldots, R^I)$, which will be constrained by a budget condition, where R^i is the amount that will be spent to treat each case of disease i, regardless of the characteristics of the person who has contracted the disease. Thus, *by definition*, we restrict ourselves to policies that are *horizontally equitable*: Any person suffering from disease i, regardless of her type and lifestyle quality, will receive the same treatment, because treatment expenditure is not a function of these variables. A more highly articulated policy space could allocate medical resources predicated also on the type of the patient and the lifestyle that patient had led. But in the health sector, doing so would set the stage for antagonistic patient-provider relations, and interfere with

[13] This point is due to Cohen (2006).

[14] Of course, persons are surely in part responsible for their socioeconomic circumstances. But the Health Ministry's mandate might be to eliminate health inequalities due to those circumstances, and so formally, it would consider socioeconomic aspects of households as circumstances.

other values we hold, and so we choose to respect horizontal equity. We will return to this point below.

For any given vector, $R=(R^1,\ldots,R^I)$ there will ensue a distribution of lifestyle quality in each type t, and a consequent distribution of disease occurrences in each type. Lifestyle quality may not be responsive to the policy, but we allow for the general case in which it is. Let us denote the fraction of individuals in type t who contract disease i when the policy is R by $p^{it}(R)$. Then the policy is *feasible* when:

$$\sum_{i,t} f^t p^{it}(R)x^i \leq \overline{R}$$

and it exhausts the budget precisely when:

$$\sum_{i,t} f^t p^{it}(R)x^i = \overline{R} \tag{4.2}$$

The set of *admissible policies* comprises all those for which (4.2) holds: This is the set Φ.

We next suppose that we know the *health production functions* for each type; these are functions that give the probability that a person of type t will contract disease i if she lives a lifestyle of quality q. Let $i=0$ represent the case of "no disease" being contracted. We denote these functions $s^{it}(\cdot)$; thus $s^{it}(q)$ is the probability that a t-type will contract disease i if she lives lifestyle quality q. We presume it is the case that $\{s^{it}\}$ are monotone decreasing functions for $i>0$; that is, raising lifestyle quality reduces the probability of disease.

We also have as data of the problem the mapping from the policy space Φ to the space of cumulative distribution functions on the nonnegative real numbers. Denote that class of distribution functions by Γ. The map

$$F^t : \Phi \rightarrow \Gamma$$

gives us the distribution of lifestyle qualities that will occur in type t, at any policy R in Φ. We write $F_R^t = F^t(R)$. Thus, an individual with lifestyle quality q in type t lies at rank π of the effort distribution of her type, when the policy is R, if $F_R^t(q)=\pi$. We denote this value of q by $q_R^t(\pi)$.

Finally, we need to postulate the relationship between treatment of disease and health outcome. Let us take the outcome to be life expectancy. We therefore suppose that we know the life expectancy for those in type t who have contracted disease i and who are treated with the resource expenditure specified by R. Denote this life expectancy by $\lambda^{it}(R)$. (Denote by λ^{0t} the life expectancy of a person of type t who contracts no disease.) We could further complexify, here, by assuming that life expectancy is a function, in addition, of the lifestyle quality of the individual, but choose not to do so.

Consider, now, a policy $R=(x^1,\ldots,x^I)$, which induces a distribution of lifestyle quality in each type. Consider a type t and all those at rank π of t's lifestyle quality distribution. Assume there is a large number of people in each type, so that the fraction of people in a type who contract a disease is equal to the probability that people in that type will contract the disease. Then,[15] the average life expectancy of all such people—the (t,π) cohort—will be

$$s^{0t}\left(q_R^t(\pi)\right)\lambda^{0t}+\sum_{i=1}^{I}s^{it}\left(q_R^t(\pi)\right)\lambda^{it}(R)\equiv L^t(\pi,R)$$

We can now define the EOp policy, which is:

$$R^{\mathrm{EOp}}=\arg\max_R\int_0^1\min_t\ L^t(\pi,R)\mathrm{d}\pi \tag{4.3}$$

Although we need a lot of data to compute the EOp policy, it is only the Ministry of Health who must have these data: once the policy is computed, a hospital need only diagnose a patient to know what treatment is appropriate (i.e., how much to spend on the case). No patient need ever be asked her type or her lifestyle characteristics. There is, that is to say, no incursion of privacy necessitated by applying the policy—apart from the initial incursion in the research survey on a population sample that assembles the data set to compute the health production functions. The policy is horizontally equitable. This is an important point, because some philosophers have falsely concluded that applying the equal-opportunity approach will necessitate incursions into privacy, and making distinctions among individuals in resource-allocation questions that are either difficult or socially objectionable in some way (see Anderson, 1999). But this is incorrect: The planner can choose the policy space in a way that makes such distinctions irrelevant for implementing the policy. In other words, not only is the delineation of circumstances a political/social decision that may vary across societies, but so must the specification of the policy space take into consideration social views concerning privacy and fairness.

Let us make this example numerical. We posit a society with two types, the rich and the poor. The poor have lifestyles whose qualities q are uniformly distributed on the interval [0,1], while the rich have lifestyle qualities that are uniformly distributed on the interval [0.5, 1.5]. The probability of contracting cancer, as a function of lifestyle quality (q) is the same for both types, and given by:

$$s^{\mathrm{CP}}(q)=s^{\mathrm{CR}}(q)=1-\frac{2q}{3}.$$

Only the poor are at a risk of tuberculosis; their probability of contracting TB is:

$$s^{\mathrm{TB}}(q)=1-\tfrac{q}{3}.$$

[15] In the formula that follows, we have assumed for the sake of simplicity that an individual contracts either no or one disease. Of course, the formula can be generalized to the case where we drop this assumption, as we do in the numerical example that follows.

Suppose that life expectancy for a rich individual is given by:

70, if cancer is not contracted, and

$60 + 10\dfrac{x_C - 1}{x_C + 1}$, if cancer is contracted, and x_C is spent on its treatment.

Thus, if the disease is contracted, life expectancy will lie between 50 and 70, depending on how much is spent on treatment (from zero to an infinite amount). This is a simple way of modeling the fact that nobody dies of cancer before age 50.

Suppose that life expectancy for a poor individual is:

70 if neither disease is contracted,

$60 + 10\dfrac{x_C - 1}{x_C + 1}$ if cancer is contracted and x_C is spent on its treatment, and

$50 + 20\dfrac{0.1x_{TB} - 1}{0.1x_{TB} + 1}$ if tuberculosis is contracted and x_{TB} is spent on its treatment.

Thus, the poor can die at age 30 if they contract TB and it is not treated. With large expenditures, a person who contracts TB can live to age 70. Furthermore, it is expensive to raise life expectancy above 30 if TB is contracted. We further assume that if a poor person contracts both cancer and TB then her life expectancy will be the minimum of the above two numbers.

Finally, assume that 25% of the population is poor and 75% is rich, and that the national health budget is $\overline{R} = \$3000$ per capita.

With these data, one can compute that 33% of the rich will contract cancer, 9.3% of the poor will contract only cancer, 26% of the poor will contract only TB, and 56% of the poor will contract both TB and cancer. (Here, we do not exclude the possibility that a person could contract both diseases.)

Our policy is $R = (x_C, x_{TB})$, the schedule of how much will be spent on treating an occurrence of each disease. The objective is to equalize opportunities, for the rich and the poor, for life expectancy.

The life expectancy of a rich person is given by:

$$L^R(\pi, x_C) = \frac{2}{3}(\pi + 0.5)70 + \left(1 - \frac{2}{3}(\pi + 0.5)\right)\left(60 + 10\frac{x_C - 1}{x_C + 1}\right)$$

and of a poor person by:

$$\begin{aligned} L^P(\pi, x_C, x_T) \;=\; & \frac{\pi}{3}\frac{2\pi}{3}70 + \frac{\pi}{3}\left(1 - \frac{2\pi}{3}\right)\left(60 + 10\frac{x_C - 1}{x_C + 1}\right) \\ & + \left(1 - \frac{\pi}{3}\right)\frac{2\pi}{3}\left(50 + 20\frac{0.1x_{TB} - 1}{0.1x_{TB} + 1}\right) \\ & + \left(1 - \frac{\pi}{3}\right)\left(1 - \frac{2\pi}{3}\right)\min\left[\left(50 + 20\frac{0.1x_{TB} - 1}{0.1x_{TB} + 1}\right), \left(60 + 10\frac{x_C - 1}{x_C + 1}\right)\right]. \end{aligned}$$

The solution of the program that maximizes the minimum life expectancy of the two types, subject to the budget constraint, is $x_C = \$686$, $x_{TB} = \$13{,}027$. In Figure 4.2, we present the life expectancies of the rich and the poor, as a function of the rank at which they sit on the effort (lifestyle) distribution of their type, at this solution. The higher curve is that of the rich. We see that, at the EOp solution, the rich still have greater life expectancy than the poor—despite the large amounts being spent on treating tuberculosis.[16] The difference, however, is much less than 1 year. Moreover, life expectancy increases with lifestyle quality—this inequality of outcome is an aspect that EOp does *not* attempt to eliminate.

Let us compare this solution to the *utilitarian* solution, the expenditure schedule at which *life expectancy in the population as a whole* is maximized. The solution turns out to be $x_C = \$1915$, $x_{TB} = \$10,571$. Three times as much is spent on cancer as in the EOp solution. Figure 4.3 graphs the life expectancy of the two types in the utilitarian solution (dashed lines) as well as the EOp solution (solid lines).

We see that the utilitarian solution narrows the life expectancy differential between the types less than the EOp solution does (although, in absolute terms, the differences are not great in this example). The EOp solution is more egalitarian, across the types, than the

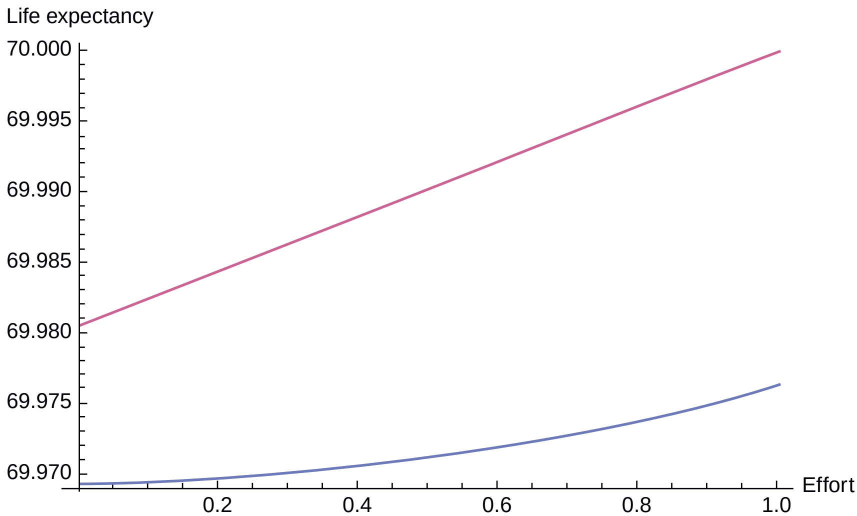


Figure 4.2 EOp policy: Life expectancy as a function of effort in two types, rich and poor.

[16] We could further reduce the difference in the life expectancies of the two types if we were willing to predicate the expenditure policy on a person's type, as well on her disease. But we have opted for a policy space that respects the social norm of horizontal equity and does not distinguish between types in the treatment of illness.

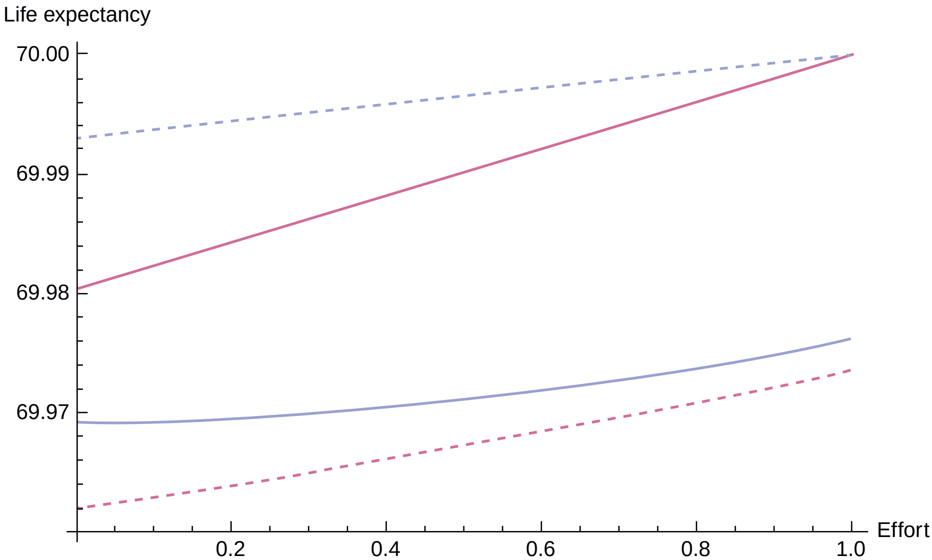


Figure 4.3 Life expectancies of rich and poor, utilitarian (dashed) and EOp (solid) policies.

utilitarian solution—the utilitarian cares only about average life expectancy in aggregate, not on the distribution of life expectancy across types.

It is obvious that different objective functions will engender different optimal solutions. The unfortunate habit that is almost ubiquitous in policy circles is to identify the utilitarian solution with the *efficient* solution. Critics of the EOp solution will say that it is *inefficient* because it delivers a lower life expectancy *on average* for the population than the utilitarian solution. But this is a confusion. Both solutions are Pareto efficient, in the sense that it is impossible, for either of them, to find a policy that weakly increases the life expectancies of everyone. Identifying the utilitarian social objective with efficiency is an unfortunate practice, rooted in the deep hold that utilitarianism has in economics. *Social* efficiency is defined with respect to whatever the social objective is, and there are many possible choices for that objective besides the social average. We discuss this point with respect to measuring economic development below in Section 4.5.

4.4. A MORE GENERAL APPROACH

Formula (4.1) gives an ordering on policies, with regard to the degree to which they equalize opportunities, after the set of circumstances has been delineated. It implements the view that inequalities due to differential circumstances for those who expend the same degree of effort are unacceptable. There is, however, a conceptual asymmetry: while the instruction to eliminate inequalities due to differential circumstances is clear, the permission to allow differential outcomes due to differential effort is imprecise. How much reward does effort merit? There is no obvious answer. To provide a social-welfare function (or a preference order over policies) that question must be answered, at least implicitly. In formula (4.1), the preference order is delineated by stating that, if there is a society with just one type, then policies will be ordered according to how large the average outcome is for that society. Fleurbaey (2008) therefore calls formula (4.1) a "utilitarian approach" to EOp.

What are the alternatives? At a policy $\varphi \in \Phi$, the *lower envelope* of the objective functions $v^t(\cdot, \varphi)$ is defined as:

$$\theta(\pi, \varphi) = \min_t v^t(\pi, \varphi). \tag{4.4}$$

We wish to render the function θ as "large" as possible: formula (4.4) measures the "size" of θ by taking its integral on [0,1]. More generally, let the set of nonnegative, weakly increasing functions on [0,1] be denoted Θ; we desire an ordering $\succeq$ on Θ which is increasing, in the sense that if $\theta(\cdot) \geq \theta^*(\cdot)$, then $\theta \succeq \theta^*$, with strict preference if $\theta(\cdot) > \theta^*(\cdot)$ on a set of positive measure. The integral of $\theta\, \mathrm{d}\pi$, as in (4.4), provides such an ordering. But many other choices are possible. For instance, consider the mapping $\Theta \to \mathbb{R}$ given by:

$$\Gamma(\theta; \varphi)\left(\int_0^1 \theta(\pi, \varphi)^p \mathrm{d}\pi\right)^{1/p} \quad \text{for } -\infty < p \leq 1. \tag{4.5}$$

Each of these provides an increasing order on Θ. As p becomes smaller, we implement more aversion to inequalities that are due to effort. As p approaches negative infinity, the order becomes the maximin order, where no reward to effort is acceptable.

We do not have a clear view about what the proper rewards to effort consist in, and hence remain agnostic on the choice of ways to order the lower envelopes $\theta(\cdot, \varphi)$. The problem of rewards to effort goes back to Aristotle, who advocated "proportionality," a view that is incoherent, as it depends on the units in which effort and outcomes are measured. Because we possess no theory of the proper rewards to effort, this is an open aspect of the theory. We believe that considerations outside the realm of EOp must be brought to bear to decide upon how much inequality with respect to differential effort is allowable. For instance, Cohen (2009) has suggested that the inequalities allowed by an equal-opportunity theory should, if they are large, be reduced by appealing to the value of social unity (what he calls "community"), which will be strained if outcome inequalities are too large.

Our agnostic view concerning the degree of reward that effort deserves contrasts with that of Fleurbaey (2008), who advocates an axiom of "natural reward" to calibrate the rewards to effort, as will be discussed in Section 4.5.

We can provide somewhat stronger foundations for the view that *an equal-opportunity ordering of policies must maximize some increasing preference order on* Θ. The first step is to note the importance of the lower-envelope function θ: for the persons who are most unfairly treated at a given policy are those, at each effort level, who experience the lowest outcomes, across types. (Hence, they are the ones represented on the lower envelope.) This is because the EOp view says outcomes that are different, due to circumstances, for those who expend the same effort, are unfair. The second step is to state an axiom which encapsulates a requirement of an EOp ordering $\succeq$ of Θ, which is:

Axiom DOM

A. *For any two policies $\varphi, \hat{\varphi} \in \Phi$, such that $\varphi \succeq \hat{\varphi}$, there exists a set of positive measure S such that $\pi \in S \Rightarrow \theta(\pi, \varphi) > \theta(\pi, \hat{\varphi})$.*

B. *For any $\varphi, \hat{\varphi} \in \Phi$ such that $\varphi \sim \hat{\varphi}$, either $\theta(\cdot, \varphi) = \theta(\cdot, \hat{\varphi})$ or there is a set of positive measure Y such that $\gamma \in Y \Rightarrow \theta(\gamma, \varphi) > \theta(\gamma, \hat{\varphi})$ and a set of positive Y' measure such that $\gamma \in Y' \Rightarrow \theta(\gamma, \varphi) < \theta(\gamma, \hat{\varphi})$.*

Part A of Axiom DOM states that if one policy is preferred to another, it must make *some* people who are among the most unfairly treated are better off than the other policy, and Part B has a similar justification. Thus, DOM is a special case of what is sometimes called the *person-respecting principle* (see Temkin, 1993): that one social alternative is better than another only if some people are better off in the first than in the second.

It is not hard to show that (see Roemer, 2012):

Proposition

Let $\succeq$ be an order on Θ satisfying DOM. Then $\succeq$ is represented by an increasing operator Γ on Θ. Furthermore, if $\succeq$ is a continuous order, then Γ can be chosen to be a continuous increasing operator.

Thus, with any continuous order on the lower-envelope functions Θ, we may write the associated EOp program as:

$$\begin{aligned} &\max \Gamma(\theta) \\ &\text{s.t.} \\ &\theta(\pi, \varphi) \equiv \min_t v^t(\pi, \varphi) \\ &\varphi \in \Phi \end{aligned} \qquad \text{(GEOp)}$$

for some increasing operator $\Gamma : \Theta \to \mathbb{R}$. The acronym GEOp stands for "generalized equality of opportunity."

We reiterate the main point of this section. Because we possess no theory of what comprise the just rewards to effort, we should not be dogmatic on the exact way to order policies. We have argued that an ordering of policies must come from an increasing order on the set of lower-envelope functions Θ, where the lower-envelope function induced by a policy φ is given by (4.5). This ambiguity in the theory results in program (GEOp), where the degree of freedom is the choice of the operator Γ. Considerations outside of the theory of equal opportunity might put constraints on the degree of overall inequality that is desirable or admissible in a society, and this can guide the choice of Γ.

We have thus argued that the theory of equal opportunity is not intended as a complete theory of distributive justice, for two reasons. First, we have emphasized its pragmatic nature. We do not have a complete theory for what people are, indeed, responsible, and have advocated the present approach as one that should be viewed as providing policy recommendations for societies that are consonant with the society's conception of responsibility. Thus, the choice of the set of types, and even of the policy space, will be dictated by social norms (we have illustrated the policy-space point with the health-expenditure example). Second, the theory does not include a view on what the proper rewards to effort consist in, and this is reflected in the openness inherent in program (GEOp).

Because we view the approach as most useful when the objective in question is something measurable like income, life expectancy, or wage-earning capacity, we shy away from taking an all-encompassing objective of "utility." We view the usefulness of the approach as one for policy makers, in particular ministries, who are concerned with narrower objectives than overall utility: the health ministry has an objective of life expectancy or infant survival, the education ministry has an objective of the secondary-school graduation rate, the labor ministry is concerned with opportunities for the formation of wage-earning capacity, or for employment, and so on. All these objectives are cardinally measurable, and it makes sense to use any of the operators defined in Equation (4.5) to generate an ordering on policies.

Nevertheless, we wish to remark that it is possible to apply the theory where the objective is "utility," if utility is cardinally measurable. (Actually, to use the operators in Equation (4.5) we require what is called cardinal measurability and ratio-scale comparability.) Because, when thinking about utility, we often conceive of effort as implying a disutility, we now show why this is not a problem for the application. Suppose utility functions over consumption and labor expended are given by $u(x, L; w)$, where $w \in W$ is the individual's wage rate. The distribution function of w in type t is given by F^t. Let us suppose we are considering the space of linear tax policies, where after-tax income is given by $(1-\varphi)wL+b$, where b is a lump-sum demogrant and $\varphi \in [0,1]$ is the tax rate. (It is implicitly assumed, since wage rates are fixed, that production is constant-returns-to-scale.) Then, the utility-maximizing individual chooses his labor supply optimally, denoted by $L(\varphi, w)$, and of course, budget-balance requires $b=\varphi \int wL(\varphi, w)\mathrm{d}F(w)$, where F is the population distribution of w. Define $w^t(\pi)$ by $F^t(w^t(\pi))=\pi$. Then the outcome functions are just the indirect utility functions:

$$v^t(\pi, \varphi) = u((1-\varphi)w^t(\pi)L(w^t(\pi), \varphi) + b, L(w^t(\pi), \varphi)),$$

and we are ready to calculate the EOp policy. Here, "effort" is interpreted not as one's labor supply, but rather as those actions the person took that gave rise to his wage-earning capacity. There are different distributions of wages in different types, reflecting the differential circumstances that impinge upon wage-formation, but within each type, there is a variation of the wage due to autonomous factors that we view as effort and worthy of reward.

4.5. THE FLEURBAEY–MANIQUET APPROACH

Marc Fleurbaey and Francois Maniquet have, in a series of writings, made a number of proposals for ordering policies with respect to the degree to which they equalize opportunities, which are similar in spirit to those discussed above, but different in detail. Their work is summarized in Fleurbaey (2008); the general inspiration of the theory is the idea of envy-freeness, pioneered in the works of Foley (1967), Kolm (1972), and Varian (1975). Here, we present one of their main proposals, which falls in the family of egalitarian-equivalent proposals, and as such, descends from the work of Pazner and Schmeidler (1978). The approach is substantially different from the one outlined in Section 4.3, because it does not take the viewpoint that equalizing opportunities involves maximizing the lower-envelope function θ defined in Equation (4.4).

Suppose that a population is characterized by an outcome function $u(c, r, \varphi)$, where c is a vector of circumstances (characteristics of the individual or his environment for which he is deemed not responsible), r is a vector of characteristics for which he is deemed responsible, and φ is a policy. We will specialize to the case where φ is the distribution of some resource to the population: say, an allocation of money. Let us suppose, further, that there is some type (i.e., vector of circumstances c^*) that characterizes the most disadvantaged type. We desire to place an ordering on policies φ that reflects the view that persons should not be held responsible for their circumstances, but should be held responsible for the choice of r.

Fleurbaey (2008) represents the idea that persons should not be held responsible for their circumstances by various "principles of compensation." An example would be "equal well-being for equal responsibility," meaning that if two individuals have the same values of r, their outcomes should be the same (i.e., independent of their circumstances). Thus, the ordering of policies should reflect this desideratum. He, Bossert (1995) and Maniquet also advocate various "principles of reward." For instance, if all individuals have identical circumstances, then the resource should be divided equally among them, called the "liberal reward principle." That is, if everyone is of the same type, there is no justification for any compensatory policy. It is clear from simple examples that it is, in general, impossible to respect the liberal reward principle and the "equal well-being for equal responsibility" principle simultaneously as long as the environment is sufficiently rich, and so Fleurbaey (2008) is a study of social-policy orderings that satisfy weaker versions of postulates inspired by these principles.

We summarize a prominent example of such an ordering. Let φ be given, and construct another allocation of the resource, $\hat{\varphi}$—which need not be feasible, given the budget—defined by:

$$u(c_i.r_i,\varphi_i) = u(c^*, r_i, \hat{\varphi}_i),$$

where i indicates the individual and c^* is a reference set of circumstances—say, those of the most disadvantaged type. Thus, under $\hat{\varphi}_i$ each individual receives an amount of resource that makes her as well off as she is in the φ allocation, but assuming, counterfactually, that she had been a member of the reference type, and had maintained the same values of the responsible factors. In the counterfactual world in which $\hat{\varphi}$ lives, everybody is of the same type (c^*) and so, *no special compensation* should be made to individuals from the opportunity-equalizing viewpoint, according to the liberal reward principle. Hence, the ideal policy φ is one in which the associated $\hat{\varphi}$ is an *equal distribution* of the resource. This tells us how to order actual policies φ: we say that $\varphi \succeq \varphi'$ if the counterfactual distribution $\hat{\varphi}$ is "more equal" than $\hat{\varphi}'$; to be precise

$$\varphi \succeq \hat{\varphi}' \Leftrightarrow \hat{\varphi} \succeq_{\text{lex}} \hat{\varphi}'$$

where $\succeq_{\text{lex}}$ is the leximin ordering.

This particular version of the egalitarian-equivalent approach to responsibility is what the authors call zero egalitarian equivalence (ZEE), because the standardization takes place by counterfactually making everyone a member of the worst-off type. Of course, standardizing with some other set of circumstances would do as well, although each choice of how to standardize will (generally) produce a different ordering over policies. One virtue of this approach is that an ordinal outcome function u is all that is required, as we only need to compare the outcome for individuals to variants of themselves (where they have different circumstances), which contrasts with the approaches discussed in Section 4.3, that require cardinality and even ratio-scale comparability.

Of course, the ZEE approach will in general give a different ordering of policies than the GEOp approach; Roemer (2012) calculates some examples. Both approaches are incomplete: GEOp, as has been discussed, does not dictate a choice of the operator Γ and ZEE does not dictate a choice of the way to standardize circumstances.

An essential feature of the egalitarian-equivalent approach is the liberal reward principle, that if everyone were of the same type, then no redistribution is called for. To be specific, in the EOp approach, Roemer closes the model by saying that if everyone is of the same type, then policies are preferred if they produce higher *average* outcomes, whereas Fleurbaey and Maniquet say that policies are better in this case the closer they are to *equal-resources*. But, as we have argued in Section 4.4, we remain agnostic on the right way of closing the model, because we do not think the concept of EOp contains a theory of just rewards to effort. In particular, the liberal reward principle, described above, will sometimes or often use market institutions to close the model. Consider a problem where all persons have the same circumstances, but preferences differ, due to voluntary choices. The principle of liberal reward might be interpreted as saying that the allocation of goods should be that associated with the competitive equilibrium following from an equal division of wealth. But this means that the welfare of individuals is determined by a particular set of institutions (markets with private property). Our objection, then, to the liberal reward principle is that in some cases there is no obvious benchmark that can be considered "natural" to define distribution in the case where there is a unique set of circumstances. This point harkens back to the legal realists, who argued that there is no conception of laissez-faire that is free of ethical bias (see Fried, 1998)—or, to put it more starkly, the usual conception of laissez-faire is a misnomer, as it presupposes property rights enforced by state power.

One disadvantage of the egalitarian-equivalent approach is that the notation does not force the practitioner to come to grips with the fact that choices people make are themselves influenced by circumstances. Recall that in the EOp approach, it was the *degree* of effort rather than the *level* of effort that was taken as reflecting responsibility, and this distinction was made because the *distribution* of levels of effort is infected with circumstances. Now one can model the same idea in the ZEE approach, but the notation does not invite doing so: There may be a tendency of practitioners to take r as *observed* levels of effort and choices of various kinds, and this would fail to take account of the fact that the distribution of choices r in a type is itself a characteristic of the type, and something that calls for compensation. So a literal application of the ZEE model, which is insensitive to this fact, will ascribe to persons responsibility for choices that are perhaps heavily influence by circumstances, and should therefore call for compensation.

One of the innovative applications of the egalitarian-equivalent approach by the authors is to tax policy. From among feasible tax policies, the policy that should be chosen is most preferred according to the ZEE preference order. As noted, this approach provides a theory of optimal taxation that does not rely on any cardinalization of the utility function. Therefore, Fleurbaey and Maniquet have produced a theory of optimal taxation liberated from cardinal measurement of utility (that is, from maximizing the integral of some social-welfare function). See Fleurbaey and Maniquet (2006) and Fleurbaey and Maniquet (2011, Chapter 11).

Fleurbaey and Maniquet also propose a kind of dual to ZEE: namely, imagine a counterfactual where all individuals expend the same reference level of effort but maintain their actual circumstances. In this case, that allocation is most preferred which most closely equalizes outcomes (that is, each person should be indifferent to how she would feel if she had the circumstances of any other person). The basis of this view is that if persons all expend the same value of the responsible factors r, then there is no ethical basis for their having different outcomes. Again, this gives a preference order on policies that can be defined without using cardinal utility functions, but using egalitarian equivalence. The authors name this approach "conditional equality."

One way to compare the approaches of Roemer and Fleurbaey–Maniquet is to ask, Can the Fleurbaey–Maniquet preference orders be rationalized as instances of program (GEOp), for some choice of Γ? It turns out that the ZEE approach can be, but the conditional-equality approach cannot be. See Roemer (2012) and Fleurbaey (2012).

Fleurbaey and Maniquet, in their work reported in Fleurbaey (2008), take an axiomatic approach, proposing a number of axioms modeling the ideas that persons should be held responsible for their autonomous actions but not for their circumstances. Strong versions of these axioms produce impossibility results, as we noted. (This is immediately clear if one thinks of the EOp model discussed in Section 4.3. There will almost never exist a policy that uses all the budget available and equalizes *for all* π, the outcomes across all types. This would be the *summum bonum*, from the viewpoint of EOp, but it cannot be achieved in a problem of any complexity. So some compromise is called for.) Their approach is to sequentially weaken axioms until they find possible preference orders over policies. A significant part of their analysis therefore consists in providing axiomatizations of different preference orders over policies, each of which has some purchase as reflecting the equal-opportunity view. The egalitarian-equivalent and conditional-equality families turn out to be the important ones.

Before concluding this section, we mention another preference ordering of policies similar in spirit to the EOp ordering, first proposed by Van de gaer (1993): order policies according to the value of

$$\min_t \int_0^1 v^t(\pi,\varphi)\mathrm{d}\pi. \tag{4.6}$$

In other words, maximize the average outcome value of the most disadvantaged type. Formally, this proposal simply commutes the integral and "min" operators compared to Roemer's approach in (4.1). Its virtue is that it is sometimes easier to compute than (4.1). If there is an unambiguously worst-off type (that is a type t such that for all policies φ and for all types t', and all $\pi \in [0,1]$ we have $v^t(\pi,\varphi) \leq v^{t'}(\pi,\varphi)$), then (4.1) and (4.6) are equivalent. Unfortunately, (4.6) is not a special case of (GEOp); it does not necessarily maximize the size of the lower-envelope function θ, for any conception of how to measure size (i.e., Γ). See Roemer (2012). Ooghe et al. (2007) compare the orderings over social policies induced by (4.6) and (4.1) by introducing a number of axioms that distinguish between the two. They argue that Roemer's approach (4.1) is a "compensating outcomes" approach, while Van de gaer's (4.3) is an "equalizing opportunity sets" approach, in the sense that the integral $\int_0^1 v^t(\pi,\varphi)\mathrm{d}\pi$ can be viewed as a measure of the degree of opportunity available to type t. Therefore, these authors link their approach to the large literature on equalizing opportunity sets (e.g., Bossert, 1997; Foster, 2011) which derived its inspiration from Sen's capability approach.

Our final topic of this section is the attempt to incorporate luck into the theory of equal opportunity. Of course, luck has already to some extent been incorporated, as circumstances are aspects of luck—for example, the luck of birth lottery assigns genes, families, and social environments. Besides the luck inherent in circumstances, however, there are two other kinds of luck that are important: first, what might be called episodic luck, which is randomly distributed across individuals, and is often unobservable to third parties (being in the right place at the right time), and the luck due to the outcome of gambles. Dworkin's view was that no compensation is due to anyone who suffers a bad outcome owing to a voluntarily taken gamble—such "option luck" is due to an exercise of preferences for which the person is held responsible. Fleurbaey (2008), however, contests this view. He splits gambles into two parts: the decision to take the gamble, which is the person's responsibility, and the outcome of the gamble, which is an aspect of luck. Let us view the risk-taking preference of the individual as a responsibility characteristic, and the outcome of the gamble as a circumstance—something over which the individual has no control. Fleurbaey proposes giving all persons with a given risk-taking propensity (i.e., responsibility characteristic) the average value of all gambles that such persons take. Thus, everyone with the same responsibility characteristic receives the same outcome. Of course, the informational requirements for implementing such a plan are severe. Moreover, this proposal seems to countervene the purpose of gambling. If gamblers wanted to protect themselves from bad outcomes, they would insure to receive the expected value of the gamble. If, however, gamblers are risk-loving, then they would only insure to receive something more than the gamble's expected value, and such insurance is not fiscally feasible. So in offering gamblers the expected value of all gambles taken by their risk type, their welfare is being reduced from actual gambling, assuming that they are true risk lovers.[17] This solution, first advocated by Le Grand (1991), has other weaknesses. The different lotteries offered to the individual decision makers can be ranked unambiguously from the most profitable to the least if Fleurbaey's solution is implemented. Indeed, the lotteries would only differ in terms of the average outcome since all risk is eliminated. All rational decision makers (who prefer more than less) will choose the same lottery. Full equality will be then observed *ex post*.

Lefranc et al. (2009) believe that the project of separating influences into circumstances and effort is too binary. They call "residual luck" a third influence, and recommend something weaker than compensation for residual luck, namely, that the correlation between such luck and circumstances be eliminated. Consider the following examples: Some people gain by the chance meeting of another person; popular views do maintain that persons with rare productive talent be specially compensated; the winnings

[17] Fleurbaey (2008, p. 162) distinguishes between risk lovers and super risk lovers. If they lose, the former regret gambling *ex post* whereas the latter do not. Following the distinction introduced by Kahneman et al. (1997) between decision utility and experience utility, preferences that might be respected are "experience' preferences, in the present case, the preferences of the super risk lover. Fleurbaey does not propose a compensation scheme for them.

of national lotteries (Belgium, France, United Kingdom) are often not taxed. The luck inherent in these examples (especially the first two) is often considered to be part of life, something that policy should not eliminate. The first example could be brute luck or due to special effort; the second example is brute luck; the third is option luck. These authors maintain that these kinds of luck should be equally distributed across types, at any given level of effort.

Suppose the income-generating process is given by:

$$y = g(c, e, l)$$

where c, e, and l are circumstances, effort, and residual luck, respectively. The distribution of income, conditional upon c and e is defined as:

$$H(y|c,e) = F_{c,e}\left(g^{-1}(y, c, e)\right)$$

where $F_{c,e}$ is the distribution of luck in the element of the population characterized by (c, e). The above-described principle says that

$$\text{for any } (c, c')H(\cdot|c,e) = H(\cdot|c',e) = K(\cdot|e).$$

This allows the distribution of virtual luck to depend on effort but not on circumstances. If all luck factors are named as circumstances, then the distribution K is simply a point mass. More generally, the support of this distribution can be as small as the decision maker wishes. It depends on her inequality aversion. The authors propose further refinements using stochastic-dominance arguments.

4.6. ECONOMIC DEVELOPMENT

The standard measure of economic development, GDP per capita, is inspired by the utilitarian ethic. If we identify utility with income, then average utilitarianism calls for maximizing average income. Hence, this conception of economic development is a corollary to an ethical view. As utilitarianism was ubiquitous in economic thinking until Rawls (1971), and continues to be extremely influential in economics after Rawls, especially in growth theory and policy analysis, it is unsurprising that our central measure of economic development has a basis in utilitarian thought.

There are various ways we might alter our measurement of economic development, based on other ethical views. Indeed, some alterations can be made within utilitarianism. By recognizing that some needs are more urgent than others, we could apply a concave transformation to income, say the logarithm, and measure economic development by $\sum \log x_i$, where x_i is income, which is ordinally equivalent to maximizing $\prod x_i$. Of course, this would place much more policy focus upon avoiding poverty, as a single very small income is socially catastrophic. Another approach, still within utilitarianism, is to include other arguments besides income in the utility function—education, health, etc.—but to take the average of an index of these goods over the nation. This is the approach of the UNDP's human development index. But if equalizing opportunities is an attractive ethic, then we should construct measures of economic development that are consonant with it. This section begins that discussion.

As a preliminary consideration, we must clear the deck of an opposing position which argues that economic development is a technical concept, not one related to social welfare. This cannot be correct. Economics is not engineering: Its goal is to maximize *social* welfare, however that may be conceived. Even for those who abjure the possibility of interpersonal comparisons, Pareto efficiency is a conception of social welfare. An economy consisting of slaves who produce, for a very small elite, huge wealth, should not be considered highly developed, no matter how refined the technology. Economic development must mean the development of human beings (some would include other sentient beings), and how to conceive of it must be corollary to a theory of the good life and good society.

If EOp is to replace utilitarianism as the ethical view of choice, then we must replace GDP per capita with some measure of opportunity equality as a measure of economic development. We will propose, here, a two-dimensional index of economic development, based upon the EOp approach. The first component of the index is the value of (4.1), and the second is a measure of the extent to which inequality in the society is due to inequality of *opportunity* (as opposed to differential effort).[18]

There are various methods for defining the second component; here is one. Suppose H is the distribution of income in the society, let H^t be the income distribution in type t, and let f^t be the frequency of type t. Then $H = \sum f^t H^t$. Let μ(resp., μ_t) be the mean of H(resp., H^t). Define the square of the coefficient of variation of H by:

$$C(H) = \frac{\mathrm{var} H}{\mu^2}.$$

Define the distribution:

$$\Phi^T(x) = \sum_{t=0}^{k} f^t \text{on the interval}\, \mu_k \le x \le \mu_{k+1}, \tag{4.7}$$

where $k = 0, \ldots, n$ and $\mu_0 = 0$ and $\mu_n = \infty$ and $f^0 \equiv 0$. Clearly the mean of Φ^T is μ. If Φ^T were the actual distribution of the objective in society, then everybody in a given type would have exactly the same value of income, equal to the mean income of that type. (The distribution function Φ^T is a step function with the same mean as H.) Were this

[18] For instance, take income as the objective, and define a typology by parental education levels.

the case, then the contribution of effort to inequality would be nil, as no variation of the objective would exist within any type. Now it is well known that we can decompose $C(H)$ as follows:

$$C(H) = C\left(\Phi^T\right) + \sum f^t (\rho^t)^2 C(H^t), \tag{4.8}$$

where $\rho^t = \frac{\mu_t}{\mu}$. Since both addends in this decomposition are positive, it is natural to interpret $C(\Phi^T)$ as a lower bound on the amount of inequality due to circumstances, and $\sum f^t (\rho^t)^2 C(H^t)$ as an upper bound on the amount of inequality due to effort. We therefore propose, as a measure of an upper bound on the *degree* of inequality due to effort the index:

$$\eta = 1 - \frac{C(\Phi^T)}{C(H)}. \tag{4.9}$$

The reason that the measure η is only an upper bound on the fraction of inequality due to effort is that circumstances continue to influence the second term in the decomposition (4.8). See Shorrocks (1980) for a characterization of all inequality indices that can be decomposed in the sense of (4.8).

Our proposal is to measure economic development by the ordered pair $d = (W^{\mathrm{EO}}, \eta)$. W^{EO} replaces GDP per capita: It is the average income of those who belong to the most disadvantaged type.[19] Thus, d presents both a level of welfare and a degree of inequality.[20]

The proposal to measure the degree of EOp using the decomposition (4.8) is not original with us. It is a special case of the inequality of opportunity ratio (IOR), defined in Ferreira and Gignoux (2011). Ferreira and Gignoux's preferred measure of inequality is not the square of the coefficient of variation but the mean logarithmic deviation (MLD). The same idea for measuring the degree of inequality due to circumstances is proposed in Checchi and Peragine (2010) as well.

In Figure 4.4, we present a graph plotting the points d for a set of European countries, where the data are taken from EU-SILC (2005) and the population of male workers is partitioned into three types, depending on the level of education of the more educated parent. (Type 1: parent completed only lower secondary; type 2: parent completed upper secondary; type 3: parent had some tertiary education.)

[19] Or, more generally, as we explained above, it is the average value of the objective of those in the population who comprise the left-hand envelope of the type distributions of the objective. Frequently, the left-hand envelope of the type-income-cdfs is the cdf of a single type.

[20] Atkinson suggests to us a simple normalized measure of the degree of equality of opportunity. Letting $\theta(\pi) = \min_t v^t(\pi)$, as in equation (4.1), define the degree of opportunity equality as $\int_0^1 \theta(\pi)\mathrm{d}\pi/\mu$, where μ is the mean value of the objective—that is, divide our measure of EOp, which is $\int_0^1 \theta(\pi)\mathrm{d}\pi$ by the mean. The advantage of the normalization is that it always gives a number between 0 and 1. If F is the population distribution function of the objective, then $\theta(\pi) \le F^{-1}(\pi)$ and so $\int_0^1 \theta(\pi)d\pi \le \int_0^1 F^{-1}(\pi)\mathrm{d}\pi = \mu$.

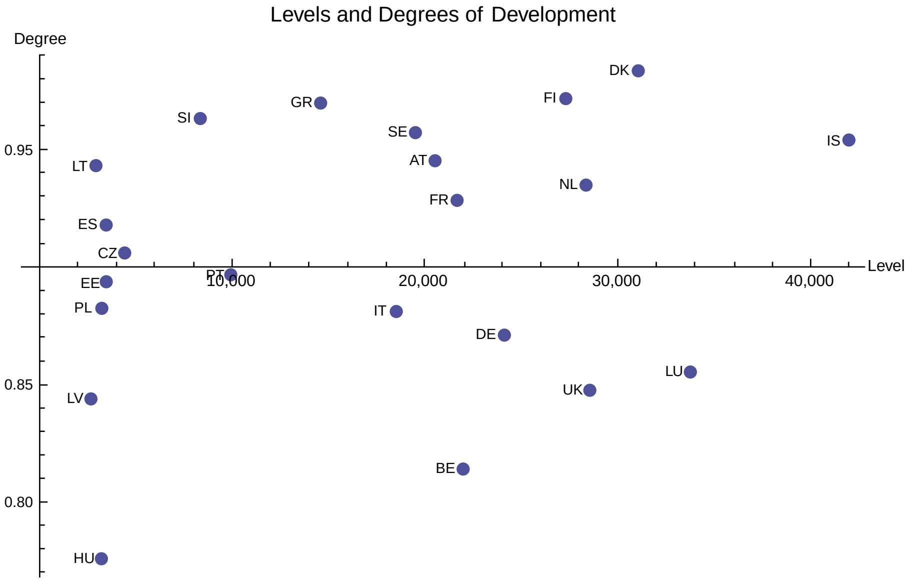


Figure 4.4 The points $d = (W^{EO}, \eta)$ for a set of European countries.

Several remarks are in order. (1) Generally, over 80% of the inequality in income is due to "effort," but recall our typology is very coarse: There is only one circumstance, parental education, partitioned into three levels. A finer decomposition of the population into more types would lower the degree of inequality due to effort. (2) Iceland's (IS) strong position on the first component, it must be remembered, is from data before the bank crisis. (3) No country dominates all others on both components of d. But Denmark (DK) dominates all other countries except Luxemburg (LU) and Iceland. (4) Greece's high component η is due to the fact that the great majority of individuals in the sample were of the least advantaged type (their parents had not completed high school). (5) The Eastern European countries (Lithuania, Lativa, Estonia, Poland, Czech Republic, and Hungary) perform relatively poorly. Finally, recall that we are looking at highly developed countries; were we to calculate the point d for developing countries, there would be a much larger spread. (For further details on this calculation, see Roemer, 2013.)

Ferreira and Gignoux (2011) calculate their version of the measure η for six Latin American countries as well. Their calculation differs from the one presented here using the EU-SILC data in two ways: They have a different set of circumstances, and they use a different measure of inequality. There is, as one might expect, a lower degree of opportunity equalization in the Latin American countries than in the European ones.

There is one study, of Sweden, in which the population of male workers was decomposed into 1152 types, based on the observation of seven circumstances (Björklund et al., 2012). These authors use a Shapley-value method to assign the degree of income inequality due to the various circumstances and to effort. For the coefficient-of-variation-squared measure, the fraction of long-run income inequality due to effort is calculated to be between 59% and 80%, considerably lower than the 96% shown in Figure 4.4. It is a testament to the degree of EOp in Sweden that, with such a fine decomposition of the working population into types, (only) between 20% and 40% of income inequality is due to circumstances.

One disadvantage of reporting the level of economic development as a two-dimensional statistic is complexity; in particular, this generates only a partial ordering of countries with respect to the degree of development. One could create a single index by aggregating as follows:

$$\hat{d}_\alpha = \left(W^{\mathrm{EO}}\right)^\alpha \eta^{1-\alpha} \tag{4.10}$$

for some $\alpha \in (0,1)$. The advantage of the Cobb-Douglas aggregation is that the ordering it imposes on countries is independent of the units in which W and η are measured, so it does not matter that W is a large number and η is a small one. For the European countries in Figure 4.4, most values of α in (0,1) render a country-ordering that is very highly correlated with the ordering of the first component. We conjecture that this would not occur with a larger set of countries, in which the variation of η would be more substantial.

The World Bank has been an important innovator in bringing considerations of equal opportunity into economic development. Its two important publications to date have been the 2006 World Development Report, *Equity and Development*, and a monograph, *Measuring Inequality of Opportunities in Latin America and the Caribbean* (Paes de Barros et al., 2009). The more recent publication contains a wealth of information on the effects of social circumstances on various measures of achievement and output.

Paes de Barros et al. (2009) propose a measure of EOp. Consider a particular kind of opportunity, such as attaining the sixth grade in elementary school. Let the total sixth-grade attendance in a country be H, and the total number of children of sixth-grade age be N, and define $\overline{p} = \frac{H}{N}$ to be the *access* on average of children to the opportunity of a sixth-grade education. $\overline{p}$ measures the level of this opportunity in the country, but not the extent to which access is unequal for different children, based on their social circumstances. Now using a logit model, they estimate the probability that each child, j, in the country has of attending the sixth grade, where that probability is a function of a vector of circumstances; denote this estimated probability by $\hat{p}_j$. Define $D = \frac{1}{2\overline{p}N}\sum\left|\hat{p}_j - \overline{p}\right|$. D measures the variation in access to the opportunity in question across children in the country. The normalization guarantees that $0 \le D \le 1$. Now define the *human opportunity index* as

$$O = \overline{p}(1 - D);$$

note that $0 \le O \le \overline{p}$.

The human opportunity index is a nonconsequentialist measure of development, because the probabilities $\hat{p}_j$ can only be computed knowing the circumstances of the children. The measure combines a concern with the level of provision of opportunities and the inequality of the distribution of them. This is to be contrasted with the ordered pair $\left(\hat{W}^{\mathrm{EO}}, \eta\right)$, which separates these two concerns into two measures. Obviously, some information is lost in using a single measure rather than two measures.

The concern of the 2009 report is in large part with children. In our view, where children are concerned, all inequality should be counted as due to circumstances, and none to effort, and so the fact that the human opportunity index does not explicitly make the distinction between effort and circumstances is unobjectionable.[21] However, if the measure is used for addressing inequality of opportunity for adults, this may be a defect.

To study this, let us take an opportunity for adults—the capacity to earn an income above M. Suppose there are three types of worker, according to the level of education of their more educated parent. Denote the distribution of income in type t as F^t; let the fraction of type t be f^t and let F be the distribution of income in the society as a whole. Then $\overline{p} = 1 - F(M)$ is the average access to the opportunity in question in the country. Now for all members j of a given type, t, compute that $\hat{p}_j = 1 - F^t(M)$: This is because the probabilities $\hat{p}_j$ are computed by taking the independent variables in the logit regression as the circumstances. Hence, the human opportunity measure is:

$$\begin{aligned} O &= \overline{p}\left(1 - \frac{1}{2\overline{p}}\sum f^t |1 - F^t(M) - (1 - F(M))|\right) \\ &= (1 - F(M)) - \frac{1}{2}\sum f^t |F(M) - F^t(M)|. \end{aligned} \tag{4.11}$$

Despite the fact that effort is not explicitly mentioned in defining the index, effort is reflected in the measure, because the distributions F^t appear in the calculation. Indeed, the first term $1 - F(M)$ measures the level of opportunity in the country, while the second term is a penalty for the degree to which this opportunity is mal-distributed with respect to circumstances (e.g., if there was no inequality of opportunity, then $F^t(M) = F(M)$ for all t, and the penalty is zero).

In expression (4.11), the first term on the right-hand side, $1 - F(M)$, plays the role that $\hat{W}^{\mathrm{EO}}$ plays in the ordered-pair measure we introduced above: It measures the level of

[21] Children should only become responsible for their actions after an "age of consent" is reached, which may vary across societies. Both the effects of nature and nurture should be considered circumstances for the child.

development. But while $\hat{W}^{\mathrm{EO}}$ focuses upon how well off the most disadvantaged type is doing, $1 - F(M)$ is a level for the society at large. The second component of our measure, η, is explicitly derived to show the degree to which inequality is due to circumstances, whereas the second term on the right-hand side of (4.11) is a form of a variance. Certainly, these two measures are getting at the same phenomenon. We have a slight preference for our proposal, as it is more carefully justified as measuring what we are concerned with. But these are minor differences; certainly, the measure O is in the spirit of thinking of economic development as opportunity equalization.

We finally consider a confusion (from our viewpoint) that infects discussions of "equity versus development," similar to the one we mentioned when we presented the health-expenditure example. It is often said that equity and efficiency are competing goals, and that equity is purchased at the expense of efficiency. There are two senses in which this phrase is uttered. The first is that redistributive taxation may be purchased only at the cost of *Pareto* inefficiency, due to workers and firms facing different effective wages. This is true. The second sense is that redistribution may lower total output. These two claims are in principle independent. There may be policies that reallocate income in a more equitable manner and lower total output, but are not Pareto inefficient. (Think, for example, of reallocating educational funds from tertiary education to secondary education in a poor country. This might have a purely redistributive effect, without significant consequences for Pareto efficiency.)

We wish to criticize the second usage of the phrase. Saying that there may be a trade-off between equity and efficiency *where efficiency is measured as total output* is equivalent to saying there is a trade-off between equity and the *utilitarian* measure of development, which (in its simplest form) is given by output per person. Consider the following quotations from the otherwise fine World Development Report 2006, issued by the World Bank, entitled *Equity and Development*. In these quotations, equity and development are counter-posed:

> *Greater equity is thus doubly good for poverty reduction: through potential beneficial effects on aggregate long-run development and through greater opportunities for poorer groups within any society (p. 2)*

> *If the opportunities faced by children like N. are so much more limited than those faced by children like P. or S., and if this hurts development progress in the aggregate, then public action has a legitimate role in seeking to broaden opportunities. (p. 3)*

> *Third, the dichotomy between policies for growth and policies specifically aimed at equity is false. (p. 10)*

In the first quotation, saying that equity is "doubly good," in that it is good for the poor and also good for long-run development, only makes sense if one assumes that equity and long-run development are *different goals*. In our view, long-run development *means* approaching equity—that is, EOp. We believe that the authors of this sentence had in mind GDP per capita as the measure of long-run development, and so what is being said is that equalizing opportunities will increase GDP per capita. This is peculiar in a report that is devoted to advocating the view that economic development requires the achievement of equal opportunity.[22] In the second quotation, the assumption is that redressing the inequality of opportunity among the children is justifiable because that inequality *hurts development*; but in our view, it is that inequality which *comprises* underdevelopment, and so the sentence is tautological. Here, the authors have in mind a utilitarian concept as the measure of economic development. Finally, the third quotation would likewise be a tautology for us; but in the context, the authors are saying that policies that increase *EOp also lead to an increase in total income*. (That is, the third quotation is offered as an empirical claim, while for us, it is a tautology.) Again, there is an ambivalence in the conceptualization of economic development: Does it mean equalizing opportunities or increasing per capita output?

It will often be the case that policies that redress inequality of opportunity will also increase total output, because improving opportunities for the disadvantaged releases talents that were, before, unused. But this need not be the case, and we maintain that our justification for redressing inequality of opportunity should not depend on its being the case. There may be groups in society that are so disadvantaged that it is very costly to compensate them: The return in output per funds invested may be small. Equity may be advanced only by shifting investment from uses where it generates high output to ones where it generates lower output. (This may be so, particularly in the short run.) But if this is the case, it does not mean that the policy in question should not be undertaken, nor does it mean that development is thereby reduced if it is.

The ambivalence in *Equity and Development* is a reflection of the competing conceptions of justice represented by utilitarianism and opportunity equalization. Utilitarianism, as we said, has a strong hold on economists. This is a holdover from an earlier period when utilitarianism was the only game in town—let us say, until John Rawls's work (Rawls, 1958, 1971). Economists and mathematicians developed optimization techniques (e.g., the Bellman equation), which are suited to solving problems where utilities are added up across persons, but not to solving problems where the minimum is maximized. And so it is often comfortable to work with utilitarian formulations. We submit, however, that this is a bad habit we should not continue to practice.

If our view of economic development is adopted, there may be a significant change in policy evaluation. One would not have to justify investment in very disadvantaged social groups by showing that such investment increases total output. As we indicated, in the long run, such a conflict might not exist, but often, policy makers are under political pressure to evaluate the consequences of their policy choices in the short run. If a country is

[22] To say that development "requires" equalizing opportunities is weaker than saying that it is synonymous with equalizing opportunities: We have been advocating the latter position in this section.

evaluated on the basis of its ordered-pair statistic $d = (W^{\mathrm{EO}}, \eta)$ rather than on GDP per capita, policies could be quite different.

4.7. DYNAMICS

EOp invites a dynamic approach. If we apply an EOp policy today, what effect will it have on the distribution of types in the next generation? One hopes that sequential application of EOp policies would create a society in which most of the effect on inequality from circumstances has been eliminated. A natural way to study this question is to analyze stationary states; that is, policies which have the property that the society they produce at date $\tau + 1$ is a replica of the society that existed at date τ.

We know of only one paper on this topic, by Roemer and Ünveren (2012), which presents an extended example. In the society postulated, there are two economic classes, rich (R) and poor (P), whose pretax (inelasticallly produced) incomes are w_R and w_P, $w_R > w_P$. Both the family and state invest in children. Let private investment in its child by a type J family be i_J and state investment in a J child be s_J, for $J \in (P, R)$. At a point in time, the fraction of $R(P)$ households is $f_R (f_P = 1 - f_R)$. Mean income at this time is $\mu = f_R w_R + f_P w_P$. The state investments are funded by a linear income tax at some rate t; thus

$$t\mu = f_R s_R + (1 - f_R) s_P. \tag{4.12}$$

Let $z_J = i_J + s_J$ be the total monetary investment in a J child, $J \in (P, R)$. The probability of the child's being successful, in the sense of becoming an R adult, is a function of his background. For a child growing up in an R household, it is

$$\pi_R(z_R, z_P) = \frac{e^{z_R}}{e^{z_R} + e^{z_P}}, \tag{4.13a}$$

while the probability of transition to the R class for a child from a P background is:

$$\pi_P(z_R, z_P) = \frac{a e^{z_P}}{e^{z_R} + e^{z_P}}, \quad 0 < a < 1. \tag{4.13b}$$

The fact that $a < 1$ models the idea that the cultural effects of growing up in a P household (and neighborhood) reduce the chances of becoming an R adult. The formulation of the transition probabilities is a reduced-form representation of a process of competition for the "good" jobs among young workers.

The *standard of living* of a J adult is his after-tax income, which is $y_J = (1 - t) w_J - i_J$. The *utility* of an adult is a function of his income and the expected income of his child when she becomes an adult. We may write the utility of a J adult at date τ as:

$$U_J^\tau = y_J + \varphi\left(\pi_J^\tau y_R^{\tau+1} + \left(1 - \pi_J^\tau\right) y_P^{\tau+1}\right). \tag{4.14}$$

A *stationary state* is a stable set of policies and decisions. It comprises a policy (t^*, s_P^*, s_R^*), optimal private-investment choices by households, (i_R^*, i_P^*), and a stable fraction of rich households f_R^*, such that the following hold:

(1) $t^*\mu^* = t^*\left(f_R^* w_R + \left(1 - f_R^*\right) w_P\right) = f_R^* s_R^* + \left(1 - f_R^*\right) s_P^*,$

(2) i_R^* maximizes (over i)

$$\left.\begin{array}{l}(1-t^*)w_R - i + \\ \varphi\left(\pi_R\left(s_R^* + i, z_P^*\right)\left((1-t^*)w_R - i_R^*\right) + \left(1 - \pi_R\left(s_R^* + i, z_P^*\right)\right)\left((1-t^*)w_P - i_P^*\right)\right)\end{array}\right\} \text{Program } P_R$$

(3) i_P^* maximizes (over i)

$$\left.\begin{array}{l}(1-t^*)w_P - i + \\ \varphi\left(\pi_P\left(z_R^*, s_P^* + i\right)\left((1-t^*)w_R - i_R^*\right) + \left(1 - \pi_P\left(z_R^*, s_P^* + i\right)\right)\left((1-t^*)w_P - i^*\right)\right)\end{array}\right\} \text{Program } P_P$$

(4) $f_R^* \pi_R\left(z_R^*, z_P^*\right) + \left(1 - f_R^*\right)\pi_P\left(z_R^*, z_P^*\right) = f_R^*$

Condition (1) is the budget constraint, and condition (4) says that the fraction of R households is stable; condition (2) defines the optimal investment choice of an R parent, knowing that the next period will look exactly like the present period from the viewpoint of his child. Condition (3) defines the optimal investment choice of a P parent in the stationary state.

Write

$$I_J = \left\{i_J \geq 0 : i_J \text{ solves Program } P_J\right\}, J = R, P.$$

An *environment* is summarized by the data (w_R, w_P, a, φ) with the intergenerational transmission functions (π_R, π_P). For this environment, there will exist a set of stationary states. We are interested in the stationary state that is best from the equal-opportunity viewpoint. We define this as follows. In a stationary state, the expected standard of living of a J child is:

$$E_J = \pi_J((1-t)w_R - i_R) + \left(1 - \pi_J\right)((1-t)w_P - i_P)$$

The equality-of-opportunity ethic maintains we should maximize the expected standard of living of the worse-off type of child. Thus, if ξ and ξ^* denote two stationary states, then EOp weakly prefers ξ to ξ^* if:

$$\min_{J=P,R} E_J(\xi) \geq \min_{J=P,R} E_J(\xi^*). \tag{4.15}$$

Obviously, the ordering on stationary states defined by (4.15) induces an ordering on policies. We wish to compute the most desirable state policy according to the preference order (4.15).

Solving for the optimal stationary state is complicated, because the optimization program is nonconvex due to the incentive-compatibility constraints. The authors compute optimal policies for a randomly generated set of economies by analysis and simulation.

The striking result is that, in 76% of the economies randomly generated, the optimal stationary state from the EOp viewpoint is *laissez-faire*; that is, the state should neither tax nor invest in children. The reason is that if the state invests in poor children, rich families compensate by investing more in their children.

Admittedly, this is just an example. The authors then consider a second type of policy: investment in parents. Formally, this is modeled by devoting state investment to raise the coefficient a (see Equation (4.13b)), which reduces the handicap that poor children face due to their background. Now, in the simulations, in 80% of the cases, the state invests in parents (that is, in increasing a), but not in children.

These results are mindful of the work of Heckman (2011), who has been championing the importance of early childhood education. It appears that much of the disadvantage of being poor has already occurred by the age of three or four. We suggest, based on these results, that investment in poor families may be more productive, in the long run, than investing directly in children.

Finally, a more radical solution to the disappointing result that rich parents will often undermine, through private investments, the effort of the state to equalize opportunities for children through educational investment, is to ban private education. This is essentially what has been done in the Nordic countries, and it is perhaps no coincidence that these countries perform among the best in the world in terms of social mobility and equalization of opportunities.

A second approach to incentive issues in EOp is the work of Calsamiglia (2009), who points out that if there are several ministries attempting to equalize opportunities for different objectives, each taking a "local" approach, the consequence may be to not equalize opportunities globally. Her paper characterizes the types of local EOp policies that will induce global EOp.

Suppose that Paul and Richard have identical preferences and skills; both want to play professional basketball and to attend college. They face the same basketball resources in their two neighborhoods, but Richard's (rich) neighborhood has better schools. So Richard is advantaged with respect to the probability of college admission due to a fortunate circumstance. Their probabilities of being admitted to college and a professional basketball team will depend on their efforts in school and in basketball, respectively, and on the resources in their neighborhoods.[23] Suppose initially that both pro-basketball and college recruiters adopt a "market" policy: they admit candidates based only on their scores on relevant tests, which are functions of effort and circumstances in the relevant arena. Facing these policies, Paul and Richard choose basketball and school effort (e_B, e_S) to maximize the total probability of admission to the basketball league and college, minus some convex cost in total effort. Since school effort is relatively less effective for

[23] We ignore American colleges' propensity to admit star basketball players, regardless of their academic accomplishment.

Paul, he devotes less effort to school than Richard and more effort to basketball. It turns out that Richard has a higher utility, although the two boys have identical preferences and skills.

Now the basketball league and college alter their policies in an attempt to equalize opportunities. Suppose that the league's policy is to admit players based only on their efforts pertaining to basketball. Then if Paul and Richard expend the same basketball effort, e_B, they will enjoy the same probability of recruitment by the league, which is locally fair, because they have the same basketball circumstances. Suppose that the college admissions officer decides to give extra points on his college-admission score to Paul as compensation for Richard's advantaged circumstances: he simply adds a lumpsum to Paul's SAT score. This is also a local EOp policy. Given these two policies, Paul and Richard will not alter their efforts, because of the lump-sum nature of the compensation to Paul, and hence Paul and Richard will have the same probability of college admission (locally EOp), but Paul has a higher probability of getting into the basketball league, as he expended more basketball effort. Although the policies are each *locally* EOp, the global result is not opportunity equalizing.

The problem lies with the lump-sum nature of the EOp policy in the college sector. Calsamiglia proves that, under assumptions that the environment is sufficiently rich, the necessary and sufficient condition for local EOp policies to aggregate to a global policy that is opportunity-equalizing is that the *marginal* returns to effort must be identical for all candidates in each sector. Because Paul's effort in school is less remunerative than Richard's, due to his inferior school, the proper policy is to augment the *returns per unit of school effort* for Paul in terms of the desired outcome (probability of college admission).

Certainly, many affirmative action policies are of the wrong, lump-sum type. For example, universities often given extra points to students from disadvantaged backgrounds, in considering admissions. The empirical implications of Calsamiglia's result have yet to be examined.

4.8. PREPARING THE GROUND FOR EMPIRICAL ANALYSIS

The literature on distributive justice is divided into two strands, a large normative one and a small descriptive one. The previous sections have considered the normative foundations of EOp. This section and the next review the empirical evidence showing that in many societies, ordinary people distinguish between two causes of inequality: those for which individuals should not be held responsible, and those for which they should be. If people do make this distinction when discussing inequality, then implementing opportunity-equalizing policies may be politically more feasible than otherwise. The issue of social acceptance of the principle is even more important if one follows Roemer's (1993) view according to which the cut between circumstances and effort should be a social and cultural decision, rather than a metaphysical one. Each society should determine the precise set of variables that describe the circumstances and the effort variables according to the views of its population. Intercultural differences in social preferences will obtain in this pragmatic view of EOp. Empirical work on intercultural differences in the attribution of the responsibility is then relevant. The state of our knowledge on these matters is still weak. Below, we list the most obvious candidates for an empirical assessment.

The first issue concerns the so-called responsibility cut. In the philosophical literature, there is a debate between those who advocate that people should be responsible for their preferences (for example, Dworkin, 1981a,b; Fleurbaey, 2008) and those who argue that the responsibility variables should be those under the control of the individual (prominently, Arneson, 1989; Cohen, 1989).

The second issue concerns the correlation between effort and circumstances. Lifestyle choices (patterns of alcohol use, exercise, smoking, diet, and so on) are examples of variables under proximate personal control. These choices are, however, influenced by family and social background. As we have said, for the measure of effort to be appropriate for the theory, it must be sterilized of the impact of circumstances upon it. "If we could somehow disembody individuals from their circumstances, then the distribution of the propensity to exert effort would be the same in every type" wrote Roemer (1998). As we wrote earlier, Roemer's technique for sterilizing effort of the effect of circumstances upon it is to measure the degree of a person's effort by her rank on the distribution of effort of those in her type. The same issue arises with preferences: If a large number of persons in a given type have preferences which, let us say, degrade the value of education, one must recognize that educational choices of such persons are influenced by their circumstances, and are not autonomous in the appropriate sense. Dworkin's (1981b) opposition to this move is to claim that *not* holding persons responsible for their preferences is to disrespect them. Another philosopher who opposes sterilizing the effort distribution of its circumstantial causes was Brian Barry, who believed that persons should be rewarded for hard work, even if that was induced by familial culture and pressure.

The responsibility cut must also to be drawn among the different kinds of luck.[24] As we wrote, Dworkin (1981b) distinguished between brute and option luck. A typical example of option luck is the outcome of a deliberate gamble. As we wrote, Fleurbaey (2008) does not advocate holding individuals responsible for the entire consequences of option luck. He attempts to disentangle the risk-taking aspect from the purely random aspect of a gamble, considering the latter to be a circumstance. Various compensation schemes respecting this distinction are proposed.

Implementing EOp may be viewed as weakening the traditional role of the family. Roemer (2004) has proposed that parents affect the opportunities of their children

[24] Alesina and Angeletos (2005) argue that societies are divided on the importance played by luck in shaping outcomes.

through four channels: (C1) the provision of resources and social connections, (C2) the formation of beliefs and skills in children through family culture and investment, (C3) genetic transmission of ability, and (C4) the formation of preferences and aspirations in children. He views the first three as circumstances, deficits in which should be compensated by an equal-opportunity policy. Preferences and aspirations are more complicated. If a coal miner loves coal-mining culture and instills in his child the desire to become a miner, this is a legitimate influence that does not call for compensation. What better conception of immortality is there than transferring one's values to one's children? If, however, the parent instills that desire because he views no other career as being available to the child, that transfer of preference is not legitimate—that is to say, preferences which are themselves induced by resource deficits comprise grounds for compensation. We know of no study that attempts to disentangle the kinds of preferences parents pass on to their children in this way.

One consequence of viewing (at least some) preference transmission to children from parents as morally legitimate is to recognize that even a perfect regime of equal opportunity should not aim at equalizing the rows of the intergenerational mobility matrix. Parents may legitimately induce differential preferences in their children, *leading to differential incomes*, even if the effects of all other circumstances were miraculously compensated for. If one does not admit this, then it is difficult to justify why we do not advocate raising children collectively. At some point, when the unacceptable differential effects of socioeconomic circumstances have been largely eliminated it will become important to address the distinction discussed with respect to channel (C4).

Finally, the importance of the *nature of the objective* must be taken into account. Three important objectives appear frequently in the empirical discussion. First, education, which takes place mainly during childhood and adolescence; second, income, which is closely related to conditions in the labor market; and third, health, which matters for a lifetime. Education is peculiar because a good part of it occurs before the "age of consent," that is, the age at which people should be held at least partially responsible for the various choices they make. Health, by many, is viewed as a right, in which matters of choice should not count. Thus, the *scope* of equal-opportunity policy may differ substantially depending upon the nature of the objective.[25]

4.9. DO PEOPLE ADVOCATE EOp? LESSONS FROM QUESTIONNAIRES AND EXPERIMENTS

The information reviewed here is derived both from the answers of respondents on questionnaires and from the actions chosen by players in laboratory or field experiments.

[25] For an early survey experiment, which shows that norms of justice differ quite radically depending upon what the *distribuendum* is, see the seminal paper of Yaari and Bar-Hillel (1984).

Questionnaires are sometimes regarded with skepticism by economists, whereas they are used extensively by psychologists and political scientists (see Chapter 13 for more methodological issues). Gaertner and Schokkaert (2012) made a plea for the use of questionnaires in the field of social choice and justice and here we build upon their reasoning. What we desire is a procedure or protocol that helps subjects to reveal their norms of distributive justice. We recognize that respondents can lie; Gaertner and Schokkaert (2012) ask why respondents would do so. In the absence of self-interest, they assert, respondents will choose to reveal their true norms. (We often assume that when an agent is indifferent between cheating and telling the truth, he will tell the truth.) The main risk with questionnaires is that respondents answer at random when the question is too complex, a difficulty of which social psychologists are well aware.

4.9.1 Questionnaire on the Empirical Validity of EOp

A first source of information is provided by value surveys conducted by polling companies or scientific associations like the World Values Survey. In our opinion, these are not fully satisfactory, because the questions remain quite vague and are not related to specific normative theories. Rather, they address the beliefs of respondents concerning the determinants of success in a given country.

Since Schokkaert and Lagrou's (1983) early work, many surveys have been conducted, most of which propose vignettes about different aspects of life in order to inquire whether individuals' opinions about justice coincide with the theoretical propositions put forward by social scientists (for references and overviews, see Gaertner and Schokkaert, 2012; Konow, 2003; Schokkaert, 1999). The literature related to our topic can be divided in two subsets. The first tests the raw idea of responsibility. The second is rooted in the theories of EOp proposed by Roemer and Fleurbaey. Konow's (1996, 2001) studies, although not anchored in a theory, introduced the distinction between discretionary and exogenous variables, which is very close to the responsibility cut as viewed by Cohen (1989), although Konow was apparently unaware of Cohen's work. A discretionary variable affects output and can be controlled or influenced by the person, while an exogenous variable can have an influence on the amount or quality of output but cannot, under normal circumstances, be influenced by personal choice. His findings (telephone interviews with a general adult population of Los Angeles and written questionnaires completed by college students) support the view that for income acquisition, variables that are deemed to be controlled by the individual are viewed as legitimate influences upon income, whereas exogenous variables are not.

Perhaps the most thorough empirical study related to the philosophical project of EOp is that of Schokkaert and Devooght (2003) (see also Schokkaert and Capeau, 1991; Schokkaert and Overlaet, 1989). First, the authors test the two principles of "full compensation" and "natural reward," which are at the heart of Fleurbaey's approach (Bossert and Fleurbaey, 1996; Fleurbaey, 1995). The principle of full compensation states that two individuals who exert the same effort should enjoy the same outcome; thus, the effect of differential circumstances is fully compensated. The principle of natural reward states that, if individuals have the same circumstances, there is no reason to transfer income between them (thus, full responsibility for effort). Second, there is an intercultural dimension in their study, as they distributed the questionnaire to first-year university students in three very different countries: Belgium (April 1996), Burkina Faso (May 1996), and Indonesia (August 1997). (See also Gaertner and Schwettmann, 2007). Finally, this study highlights whether views of responsibility are sensitive to what we have defined as the objective (or the opportunity equalidandum), as the questionnaire addresses views of responsibility with respect to income acquisition and health.

Four situations are contrasted in a two-person society. The two persons differ in only one characteristic. Possibilities of redistribution between the persons are then offered, and students are asked to choose what they think is the fair *ex post* tax income distribution.

The first vignette describes a difference in preferences in income-leisure space. No explanation is offered to explain this difference in tastes, whereas the second vignette stipulates that this difference comes from different backgrounds. That vignette tests the disagreement between Roemer and Barry about sterilizing the distribution of effort of the influence of circumstances. It is important here to notice that the issue raised is not the transmission of wealth, or social networks, but the transmission of values and preferences across different generations. People convinced by Roemer's reasoning should be more inclined to redistribute from hard-working Elizabeth to easy-going Catherine in the second situation than in the first. The third and fourth vignettes concern differences in productivity. In the third vignette, the difference originates in a difference of effort in the past. The fourth vignette describes a difference in innate talent.

The results are instructive and we will present them in terms of how the majority voted. The Belgian sample made the most clear-cut choice: A majority vote for no compensation at all (no redistribution) in case of Vignettes 1, 2, and 3, and for full compensation for the situation described in Vignette 4. Thus, the Belgians endorse the view that preference for leisure is a responsibility variable—they agree with Brian Barry not to take the causal relationship with parents' preferences into account. Innate talent, however, is considered as a circumstance. Were that vote representative of Belgian choices as a citizenry, this society would possess the basic ingredients to implement an equal-opportunity policy.

The authors find that the intercultural differences are much less pronounced than one might have thought. Still, they cannot be completely ignored, since, according to the majority vote criterion, the Burkina Faso sample is indecisive for all four vignettes. The Indonesian vote is closer to the Belgian one. Indonesians share the same views on the three first vignettes, but no majority is found on the last issue, although full compensation for talent has a plurality of votes.

At this stage, it is useful to ask whether the objective matters. Schokkaert and Devooght (2003) attempted to adapt their questionnaire to health care situations. From the start, two differences with income scenarios must be noticed that render the comparison less than clear-cut. In the income case, the stakes belong to the domain of gains, whereas they belong to the domain of losses in the health care case: the health vignettes describe illness and how to cope with health care expenditures. Since the work of Tversky and Kahneman (1991), we know a person's tendency strongly to prefer avoiding losses to acquiring gains. This may explain a stronger inequality aversion in the health vignettes. In addition, if questions are asked about how to allocate a budget between two sick persons, an efficiency issue is raised, which makes it difficult to deduce views about fairness. All studies about fairness in health care (Dolan and Tsuchiya, 2009; Ubel et al., 1999, and the above-cited paper) have chosen to formulate the vignettes in a scarcity context. Of course, scarcity of resources is an important issue in the health domain (as in others), but a sequential approach with two steps might better elicit preferences about the responsibility cut.

As an example consider two of the four vignettes proposed by Schokkaert and Devooght (2003), concerning Luke and Mark who are both suffering from lung cancer. They have the same wealth at their disposal and earn the same income. Luke and Mark have to be admitted to a hospital for treatment. It is supposed that all treatments are effective. The two vignettes raise the relevance of factors that are under the control (smoking) or beyond the control (genetic) of the individual for covering lung cancer expenditure. The respondents have the choice between different divisions of the amount of public resources: equal split between the two patients, all resources for the extra cost of treating Mark, and intermediate solutions between these two.

It is noteworthy that in all three societies, equal-split garners a majority of votes in Vignette 1. A majority favor an intermediate solution when genetics calls for extra cost. The social policy that this study suggests is clear-cut: smokers should purchase private insurance for coverage of smoking-related illness. This conclusion holds as long as the society is able to attribute the cause of the extra cost to lifestyle. These results suggest that the reason that the welfare state in many countries does not appear to be inspired by responsibility-sensitive egalitarianism is not due to popular ethics, but to the difficulty of identifying an indisputable causal link in health matters. Off-piste skiing is "the exception which proves the rule," where the cost of an accident is generally borne by the individual. One salient issue remains unsettled: we know of no questionnaire focusing on the link between lifestyle and family background. The difference of opinion between Roemer and Barry has not been reflected in the empirical literature on fairness in health.

Education is another domain where we can conjecture a different attitude with respect to responsibility. Primary and secondary education take place when the person is still, arguably, below the age of consent. Richard Arneson (1990, p. 179) has appealed to this fact in egalitarian debates. Lu et al. (2013) have investigated whether primary education elicits different responses from income acquisition in the degree to which persons are held responsible for outcomes. They contrast the results obtained with two vignettes.

In the sales vignette, there are salespersons whose sales compensation is composed of two parts: a salary and a bonus. The issue concerns the fairness of the bonus. Sales depend on characteristics which are described as follows. The salesperson's circumstances are identified with his parents' network of acquaintances. Effort is described as the salesperson's hard work, and talent is described as the salesperson's skill. A salesperson's brute luck is defined by the territory to which she is randomly assigned. Finally, option luck is described as the risks the salesperson takes: he has to choose between selling an old product that has been on the market for a long time and is familiar to customers, or a more recent product with unknown customer reaction. If a bonus is to be paid to the successful salesperson, respondents are asked how fair it is to judge the salesperson by her circumstances, effort, talent, brute luck, or option luck. The respondent has to choose exactly one answer among very unfair, rather unfair, quite fair, or absolutely fair for each of these choices.

In the school vignette, pupils face difficulties at school. Remedial tuition is supposed to help schoolwork. Five factors are related to school difficulties. Circumstances are determined by parents' ability to help children with their homework. Effort is identified as the zeal with which the child does his or her homework. Talent is defined as cognitive ability, which is precisely described as an ability to concentrate. Brute luck occurs when the child missed part of the previous school year because of illness. Finally, option luck is risk taking. The child wants to be in the advanced class, with friends, but cannot keep up with the class. Respondents were asked to judge the fairness of remedial tuition, if it were necessary because of circumstances, effort, talent, brute luck, or option luck.

Figure 4.5 presents the differences in the answers to both vignettes (432 respondents in Marseilles). In the sales vignette, we interpret the answers "quite fair" or "absolutely fair" as indicating that the respondent holds the salesperson responsible for the factor. In the school vignette, we interpret the answers "very unfair" or "rather unfair" as revealing that the pupil was deemed responsible for the factor by the respondent. A chi-square test for goodness of fit is used to test whether subjects treated each factor similarly in the two vignettes. Respondents evaluated moral responsibility with respect to all causal factors except circumstances differently in the two vignettes. More specifically, salespersons were held responsible for talent, while almost no subjects held pupils responsible for talent. Only a small minority deem students responsible for risk taking, while almost everyone deem the opposite for salespersons. The difference for effort is less impressive, since a small majority of respondents still agree to hold schoolchildren responsible for their effort in doing homework. Our results are preliminary as they are perhaps influenced by framing. Nevertheless, they cast doubt on holding children responsible for educational outcomes, at least at the primary level. If that decision is implemented, then primary-school

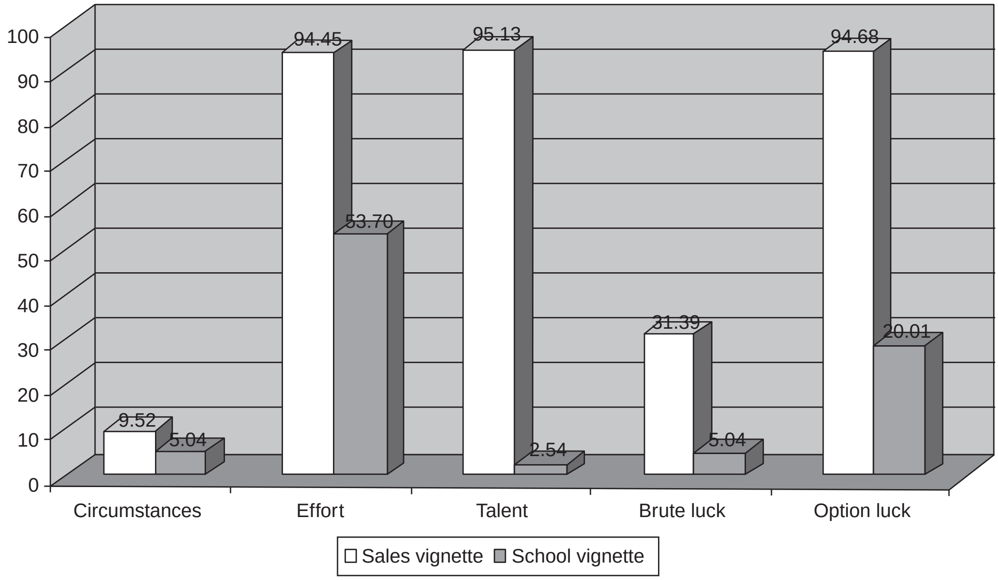


Figure 4.5 The fraction of subjects holding the agent responsible for each factor. *Source: Lu et al. (2013).*

achievements should be treated as a circumstance in studying opportunity equalization of outcomes in later life.

4.9.2 Experiments

Fairness attitudes in sharing a cake have been studied in laboratory experiments with the ultimatum game and the dictator game (Camerer, 2003), which provide a neat elicitation of preferences. These experiments reproduce exchange or distribution economies where resources are manna from heaven. Various authors (Almås et al., 2010; Cappelen et al., 2007, 2010, 2013; Frohlich et al., 1987, 2004; Konow, 2000; Rutström and Williams, 2000) have conducted experiments to study explicitly what happens to people's distributive preferences by introducing an earned-money or production stage prior to a distribution phase consisting of a dictator game. The most recent articles test the prevalence of responsibility egalitarianism among distributive justice theories. More explicitly, they investigate the control view of responsibility advocated by Cohen, summarized by the principle that "only inequalities that arise from factors under individual control should be accepted."[26]

Cappelen et al. (2007) study a situation in which individuals differ with respect both to their investments and to the rates of return that they enjoy. The agent chooses the

[26] Cappelen et al. (2007), p. 818.

amount to be invested, while the rate of return is assigned randomly. The former factor is clearly an effort variable, while the rate of return is brute luck, like talent. They assume that an individual endorses either strict equality of earnings, laissez-faire, libertarianism (each keeps his income), or responsibility egalitarianism, in which case total income is shared in proportion to investments. The distribution phase is a two-person setting in a one-shot dictator game. A parametric utility function is a weighted sum of a purely selfish element, and an altruistic quadratic loss term, which is larger, the more the distribution differs from the ideal distribution according to the individual's ethical view. The econometric analysis attempts to retrieve the parameters of the utility function, the marginal utility of money, and the preferred distributive ethic view of the subject. The authors deduce that 43.5% of subjects are strict egalitarians, 38.1% are responsibility egalitarians, and 18.4% are libertarians. The subject pool consisted of approximately one hundred students at the Norwegian School of Economics and Business Administration (NHH), a sample that cannot be viewed as representative of the Norwegian society. In addition, the results may depend on the specific form of the utility function, which balances self-interest and fairness. Nevertheless, their results confirm that responsibility-sensitive egalitarianism is endorsed by a fraction of the population and competes with libertarianism and outcome egalitarianism. But we do not learn much about the responsibility cut.

In a companion paper, Cappelen et al. (2010) use the same methodology and pool of students to enlarge the set of proposed fairness views. Individuals now differ with respect to three characteristics: working time, productivity, and the market price of their product. Subjects choose their working time (effort), market price is set randomly (brute luck), and productivity (talent) is determined through a test in the experiment (the number of correctly typed words in a short period). The authors consider four competing distributional views expressed by the list of responsibility factors. An empty list corresponds to outcome egalitarianism. If effort is the only factor belonging to this list, the view is control-responsibility egalitarianism. When this list comprises effort and talent, the view is named meritocratic[27] by the authors. (In other words, people may rightfully benefit from their inborn talent.) Finally when this list comprises effort, talent, and brute luck, it is said that the participant endorses the libertarian view. The subject pool includes students from all undergraduate years and some alumni. The differences in preferred distributive views, as estimated by the econometric model, are not pronounced among students, but alumni have quite different ethical preferences. Whatever the age group, the meritocratic view is the most popular view among students whereas the libertarian view is slightly more popular among alumni. The striking fact is that the control view of EOp is only supported by a tiny fraction of the pool: 6% among students and 2% among alumni. At this stage, it is premature to declare that these results are biased by a selection

[27] See Arrow et al. (2000) for a discussion of meritocratic ideas.

effect: however, let us remark that business-school students and alumni are very likely among the least egalitarian people in society.

In a less sophisticated way but using the same framework, Almås et al. (2010) investigate how the views about distributive justice evolve as pupils mature between the 5th and 13th grades. At the beginning of this span, schoolboys favor outcome egalitarianism (two-thirds) and libertarianism (one-third). As the children get older, they become increasingly sensitive to equality-of-opportunity arguments and by the end of the grade span, meritocracy[28] becomes the plurality view, even if it does not garner a majority of votes. Indeed it is striking that the distribution of views in this study for the 13th grade is almost the same as that obtained for the first year of college obtained by Cappelen et al. (2010).

If we assemble the lessons of these two instructive studies, they lead to the following conjecture for the development of distributive ideals over the life cycle. Starting with the stark and simple views of outcome egalitarianism and libertarianism in childhood, the development of cognitive skills induces understanding of more complex and less clear-cut views, like EOp. Views appear not to change significantly between the end of the high school and the end of the university.

Those successful in the labor market tend more toward laissez-faire opinions. Were that true in the real world, we should observe a self-serving bias (Messick and Sentis, 1983) on a large scale, in the sense that individuals, given their degree of success, would (tend to) endorse the fairness ideal that most benefits themselves. In that sense, experiments are superior to surveys and vignettes in that they enable one to measure the extent of this self-serving bias. This phenomenon should be at its minimum when subjects are students. At this stage of development, subjects are able to understand all theories of justice but they are still shielded by a veil of ignorance regarding their degree of success (in the United States, where 50% of a generation enrolls in tertiary education). The prediction would be that the difference between surveys and experiments would be minimal for this adult group.

We turn now to testing popular views about option luck. Buchanan (1986) identifies four factors that determine the distribution of income and wealth: luck, choice, effort, and birth. He considers the acceptability of rewarding effort the least controversial, and believes that the only inequalities that conflict with common views of justice are ones caused by birth (pp. 129–30). The difficulty with option luck comes from the fact that it is a mix of two more fundamental factors, one for which we want to hold people responsible, choice, and the other that is exogenous, luck. A similar difficulty prevails for talent which is a mix of birth, an exogenous factor, and past effort, which is a responsibility variable. (Buchanan does not observe the semantic convention that talent is an inborn factor, and skill results from the application of effort to talent.)

[28] This study does not make the distinction between control-responsibility egalitarianism and meritocracy.

Two papers, Cappelen et al. (2013) and Chanel et al. (2013), investigate the views of people about option luck and risk taking vis-à-vis the responsibility cut. The first article endeavors to shed light on the relative popularity of three views about option luck. The first view is Dworkin's, according to which no redistribution of gains or losses from risk taking is ethically required. Dworkin argues in favor of a laissez-faire stance, because risky lifestyles or risk taking are expressions of preferences. The second view considers it fair to eliminate all inequalities resulting from risk taking. The third view is intermediate between the first two: It would approve *ex post* redistribution between lucky and unlucky gamblers but not between gamblers and nongamblers. This view is reminiscent of a position first defended by Le Grand (1991) and refined by Fleurbaey (2008), who considers that people should be fully insured and only bear the consequences of their decisions over the expected value of the lottery. Gamblers will then receive the expected gain corresponding to their class of risk. The experiment consists of a risk-taking phase followed by a distribution phase. In the risk-taking phase, subjects face a sequence of choices between a risky and a safe alternative, where the value of the safe alternative varies. Estimates of the choice model reveal that subjects (students at the Norwegian School of Business in Bergen) have diverse opinions and split quite evenly into three groups. Roughly speaking, two-thirds of the subject pool think that people should be deemed responsible for their choice of risk taking. The same proportion but not the same individuals think that people should not bear the consequences of luck. If we interpret the econometric results as a vote, Le Grand–Fleurbaey's view is the Condorcet winner among the three alternatives offered to participants. This interesting result needs to be confirmed by other studies.

Chanel et al. (2013) are less precise in studying option luck but their aim is to deduce the relative importance of option luck in the set of factors for which individuals should be held responsible. They conduct an experiment on a large scale whose purpose is to reveal the preferences of agents when four factors matter for earnings: circumstances, effort, brute luck, and option luck. Three experimental sessions were organized involving a treatment of about 100 subjects each, who are told that they form a small society. Each treatment involves an earned-money phase followed up by a redistribution phase, where the allocation rule is determined by majority vote. In the first phase, participants can earn money through four different channels, each of which reflects a specific factor: the place of one's birth represents a circumstance and success at a visual-spatial attention task requires effort. Brute luck and option luck are easily contrasted by a random draw and taking a bet, respectively. Votes are then organized on whether or not to redistribute the gains from each step, which corresponds to a given factor. A self-serving vote is found to be prevalent (about one-third of the sample who succeeded in earning money vote not to redistribute) and nonparametric econometrics are mobilized to retrieve the true ethical preferences beneath the votes. The distribution of ethical preferences among the subject pool is described in Figure 4.6.

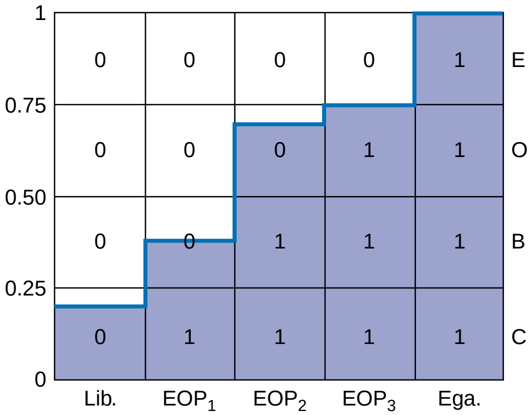


Figure 4.6 Distribution of ethical preferences about the responsibility cut. On the left vertical axis, the figures are proportions. On the right vertical axis, E stands for effort, O for option luck, B for Brute luck, C for circumstances. In each square, 0 (respectively 1) means no compensation (resp. compensation). For example, egalitarians think redistribution is mandated regardless of the cause of earnings. *Source: Chanel et al. (2013)*.

Five ethical positions are represented here.[29] At the two extremes, we find the libertarian and outcome-egalitarian stances. Three intermediate positions are allowed: in EOP1, only differential circumstances merit compensation; in EOP2 brute luck in addition merits compensation. Option luck joins the compensation set with EOP3. The two extreme positions attract almost a quarter of the views. This means that 60% of the sample endorse some version of EOp. There remains a large diversity of opinion regarding the locus of the responsibility cut. In the aggregate, the result of this experiment supports Dworkin's view according to which we should draw a distinction between option luck and brute luck, option luck being on the responsibility side along with effort, and brute luck being on the compensation side with circumstances. Nevertheless, we need to be more careful before a more definitive conclusion is reached, for many areas of uncertainty must addressed. More specifically, the design of the experiment tests Le Grand–Fleurbay's position against that of Dworkin. Redistributing gains from bettors to nonbettors has not been proposed to voters.

4.9.3 A Progress Report

In agreement with Roemer's suggestion (1993), we have developed the view that theory and empirical work are more complements than substitutes. As stated by Gaertner and Schokkaert (2012), "The theory of EOp offers a general and consistent framework which

[29] Fewer than 10% of the subjects convey an ethical preference that is not captured by one of these.

can be applied for any cut between effort and circumstances, while empirical work supplies the necessary information about where the boundary is drawn in different societies."

If we take again the four "primary factors" identified by Buchanan—birth, luck, choice, and effort[30]—it seems indisputable that subjects make a clear distinction between the first two and the last two. In questionnaire-experiments, the assumption that choice and effort are under the control of the individuals and that participants are well-informed about the consequences of the acts cannot be disputed, since the protocols of the experiments are clear. Even if more research is welcome, the conclusion reached by Konow (2001) 10 years ago appears to stand: "To summarize, the evidence from experiments and surveys generally indicates that someone whose contribution is more highly valued is more deserving if that person bears responsibility for the contribution but not if it is due to factors outside his or her control." Does this mean that from an empirical perspective, the control view of Arneson and Cohen prevails over the preference view of Dworkin and Fleurbaey–Maniquet? Not exactly, for the proper test has not been conducted. Except for Schokkaert and Devooght (2003), we know of no study testing both theories in a competitive way through questionnaire-experiments. The control theory has been repeatedly tested by psychologists and economists but not against the preference theory. We observe choices, not preferences. Economists are keen on promoting the concept of preference among social scientists; the main weakness of the concept is that preferences are not easily revealed to experts, let alone laypersons. It is asking a lot to make preferences pivotal in a theory of distributive justice that will garner mass agreement, when, at best, only some experts can argue that they have been able to deduce what preferences people hold.

EOp involves an *equalizing* aspect and a *disequalizing* one.[31] Equalization, or compensation, takes place with respect to those factors deemed circumstances; inequality is non-compensable, however, if it is due, tautologically, to factors for which individuals are held responsible. The difficulties arise when some causes of success or failure, with respect to a desirable objective, involve mixtures of these two kinds of element. Skill is a mixture of talent, due to birth, and past effort; option luck is a mixture of choice and luck. Self-protection as defined by Ehrlich and Becker (1972) is an expenditure that reduces the probability of a loss, which can be generalized to any effort that transforms the probability distribution of states in a good way for the agent. We do not know whether the

[30] One wonders why it is important to distinguish between effort and choice. An answer is suggested by Cohen, who distinguishes difficulty from costliness. It is difficult to lift a weight, but not costly; it is costly to sign a large check, but not difficult. Effort is difficult. Choice is often costly (as in taking a bet) but not difficult in the natural sense of the word. Barry's view that effort deserves remuneration even if not due to the person's choice can be explained if one believes that difficult actions deserve reward, regardless of the intent of the actor.

[31] No empirical study has tested whether people support the liberal or the utilitarian approach to reward (as far as we know).

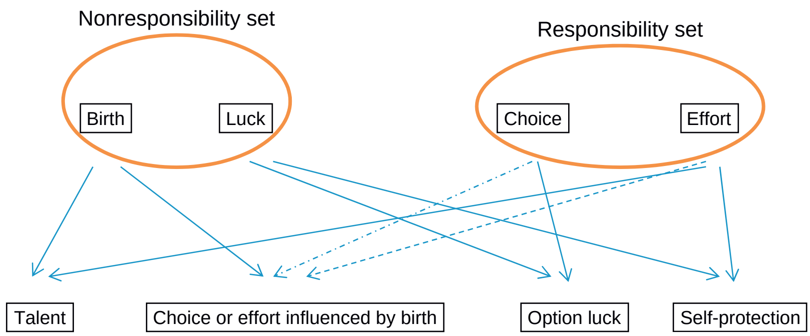


Figure 4.7 Binary combination of primary factors.

differences in views that people hold about distributive justice are due to the ambiguities introduced by the mixtures of these two kinds of factor in real life, or to fundamentally different ethical principles. See Figure 4.7.

4.10. INEQUALITY OF OPPORTUNITY: MEASUREMENT ISSUES AND EMPIRICAL RESULTS

This section will focus on methodological issues and applications of the theory. An excellent survey of the material covered in this section is provided in Ramos and Van de gaer (2012).[32]

4.10.1 Methodological Issues: General Remarks

We begin with some general remarks for the reader who is familiar with the literature on the measurement of inequality of outcomes. Measuring inequality of opportunity may mean different things. At the most basic level, we may want to encapsulate the inequality of opportunity with an index, as has been done for inequality of outcomes with the Gini, Atkinson, Theil, and others indices. We may be more modest in just wanting to rank distributions, and be content with incomplete but robust rankings provided by instruments of a dominance analysis, such as the Lorenz curve. Circumstances, effort, and luck are just sources of outcome inequality, and we may wish to trace their contribution to overall inequality. Decomposition exercises among sources are just as appropriate in EOp empirics as in inequality-of-outcome analysis. Quantifying, ranking, and decomposing are three familiar operations which we may apply to equal-opportunity analysis, and the tools are mainly borrowed from the measurement of inequality literature.

[32] See also Pignataro (2012).

4.10.1.1 EOp Measurement as a Multidimensional Problem

Nevertheless, it seems fair to say that the level of complexity of the analysis is greater because EOp is multidimensional. Equality-of-opportunity analysis may use the conceptual framework developed by Atkinson and Bourguignon (1987) in the field of multidimensional inequality. These authors focus on how to measure income inequality when each income unit belongs to a specific needs group. The information is two-dimensional—income and needs for each household—and the aim of the analysis is to rank income distributions taking into account the information provided by the vector of needs. In EOp analysis, we would rank outcome distributions (income, health, education) which are unidimensional, taking into account the information provided by the vector of circumstances, the vector of efforts and perhaps the vector of residuals. EOp measurement then belongs to the family of problems of multidimensional inequality when *margins* are fixed, where margins comprise the non-outcome information that matters in EOp assessment (circumstances, effort and perhaps the residual). The inequality in the objective must be assessed conditional on the types and efforts of the population.

A direct application of the sequential Lorenz quasi-ordering to this setting is not appropriate and it is interesting to see why. Of course, effort can be seen as analytically similar to needs; that is, at the margin, the more effort one expends, the more one deserves. Reciprocally, circumstances can be seen as negative needs: the better one's circumstances are the less one deserves. But these two statements have limitations. We may wish not to reward effort excessively, for reasons discussed in Section 4.4. And regarding circumstances, there is an asymmetry: we desire to compensate for disadvantageous circumstances, but do not regard advantaged circumstances as an evil. Furthermore it is the interplay between circumstances and effort that makes the evaluation of the ensuing inequality problematic. We need to know how additional effort should be rewarded across the circumstance dimension; as we discussed, there is no clear answer to this question within the theory. For further discussion, see Bossert (1995), Fleurbaey (1995), and Fleurbaey and Peragine (2013).

4.10.1.2 EOp as a Process

What also distinguishes EOp empirical analysis from inequality-of-outcome analysis is its two-stage nature: one generally requires an econometric-estimation stage, preceding the inequality-measurement stage. It is not so much the difference in circumstances *per se* that matters, but the difference in the impact of circumstances. Socioeconomic advantage has to be estimated through parametric and nonparametric estimation techniques, captured by the coefficient of the circumstance variable in a linear model regressing the outcome on a set of circumstances and effort variables. An evaluation of inequality must be concerned with the process that generates it. This leads Fleurbaey and Schokkaert (2009) to state, provocatively, that any EOp empirical analysis must be preceded by an estimation phase to discover the best structural model leading to the results. Only in the second step should we be interested in measuring inequality of opportunity as such.

In principle, we agree. This is, however, more easily said than done. Two observations are in order. The two main obstacles to any causal inquiry are reverse causality and endogeneity due to omitted variables. The good news is that, regarding circumstances, reverse causality can often be dismissed since circumstances are frequently characteristics of states that existed in the past (e.g., one's parents' education). However, endogeneity cannot be discarded in that way since EOp measurement is plagued with informational problems. Omitted variables are widespread; a good example is provided by genetic variables which have been found paramount in income attainment by Björklund et al. (2012). Omitted variables in empirical EOp analysis cause skepticism in claims of causality we may wish to assert. The situation is even worse when the objective is earnings, since according to Bourguignon et al. (2007), "an instrumental variable strategy is unlikely to succeed, since it is difficult to conceive of correlates of the circumstance variables that would not themselves have any direct influence on earnings." Experiments and quasi-experiments enable one to make causal statements, but experiments can usually only study problems which are much more circumscribed than those which interest researchers in this field. We are trying to understand the whole process by which someone reaches an income level, a health status, or an educational attainment. The processes are dynamic and cover part of the life span of an individual, and understanding them fully in a causal way seems out of reach at present.

Should we worry about this lack of causal interpretation? Of course, if we want to give advice to policy makers about the true effect of level-the-playing-field policies, impact evaluation needs to be causal. However, if one merely wants to measure the degree of inequality of opportunity—that is inequality due to circumstances—a correlation (with variables which occurred in the past) is already something that is relevant.

The challenge is even greater if we use the preference view for responsibility variables advocated by Dworkin and Fleurbaey. Retrieving the true parameter of the preferences is perhaps the most difficult issue in econometrics in terms of identification conditions. See, however, Fleurbaey et al. (2013) for an attempt to estimate the individual's trade-off between health and income and Bargain et al. (2013) for the estimation of cross-country preference heterogeneity in the consumption-leisure trade-off.

4.10.1.3 Lack of Relevant Information

It should be clear from this discussion that we need a much richer database to perform EOp empirical analysis than a pure inequality-of-outcome analysis. We should have variables describing the situation of the family and social background and variables pertaining to effort. It is quite common that some important background variables are missing and then we have an incomplete description of the circumstances. More importantly, effort variables are generally missing for the very reason that effort is private information, as is emphasized in economic theory. We must use proxies, which are problematical.

The measurement of effort depends on our view of responsibility. On the one hand, there is the view that effort takes into account what set of actions a person can *access*, where access is a question not simply of physical constraints, but of psychological ones, which may be determined by one's circumstances. On the other hand, there is the view that a person should be held responsible for his preferences, and hence a person is responsible for taking those actions that flow from his preferences. Roemer's measurement of effort as the rank of a person's effort in the distribution of effort of her type represents the access (or control) view: one judges the accessibility of actions to members of a type by what people in that type actually do. (This view is also reflected in Cohen's, 1989 phrase "access to advantage," which he desires to equalize.) Dworkin and Fleurbaey represent the preference view, in which a person is held responsible for his choices, if they flow from preferences with which he identifies. Because almost all empirical studies (except Fleurbaey et al., 2013; García-Gómez et al., 2012) seem implicitly guided by the control view, the authors should explain in what sense the chosen variables are under the control of the individual. Jusot et al. (2013) have argued that lifestyles in health (diet, exercise) are examples of variables under the control of the individual, and inequality of opportunity for achieving health status should be measured with this in mind.[33]

Several points that should be made about two variables that appear repeatedly in empirical analysis when trying to measure EOp in income attainment: the number of hours of work and years of education. The number of hours of work is a good effort variable, under the control view, for self-employed occupations, but is clearly less satisfactory for wage earners. It is true that hours of work correspond to a quantum of effort: The issue is whether they correspond to the *desired* amount of hours. Part-time jobs may be involuntary; overtime work may depend on the orders of the firm, and obviously unemployment may be just bad luck. To a large extent, using hours of work in a given period as an effort variable is therefore problematical for wage earners. We can be more confident that the number of hours of work over the life span is under the control of the individual because one can compensate for the impact of bad luck and low hours of work during a given period by working more in luckier periods. Using the full data for the life span is, however, quite rare (See Aaberge et al., 2011 or Björklund et al., 2012 for examples.) For snapshot distributions, the question arises of how to purge hours of work of bad luck, which, by assumption is not under control of the individual. Detecting chosen part-time from involuntary part-time is a difficult econometric issue. At best, we would

[33] See Rosa-Dias and Jones (2007), Rosa-Dias (2009, 2010), García-Gómez et al. (2012), and Van de gaer et al. (2012) for papers measuring inequality of opportunity in health.

estimate a probability that the person works voluntarily part-time, which makes the effort variable a number in the interval [0, 1]. Any empirical study that fails to do so will not respect Fleurbaey and Schokkaert's methodological dictum to do the best to estimate the most thorough structural model before any attempt is made to measure inequality of opportunity.

Years of education is also a popular effort variable in empirical studies. It is controversial to consider it as a variable under individual control, because primary and secondary education take place when the person is a child and adolescent, largely prior to the relevant age of consent. If a child is lazy in school, there might be factors not under his control that explain his laziness. Only tertiary education and lifelong learning are immune from this criticism. The problem with tertiary education comes from its path-dependency: One's probability of being accepted to university depends on one's grades in secondary education, which, in turn, depend on achievements in primary school. The above-mentioned problem for the two early stages of education then contaminates higher education attainment.

A good starting point is to attempt to account for achievements in early education by circumstances of the family. Socioeconomic circumstances may be available in data sets, but parental pressure to achieve is also an important determinant of educational outcomes, and is usually not measured. We cannot, therefore, usually give a complete account of educational achievement. However, if one views all actions of the child as due to either nature or nurture, both of which are beyond his or her control, by hypothesis, before the age of consent, then one should simply take the child's educational accomplishments at the age of consent as a circumstance with respect to determining outcomes in later life. Family circumstances may still be important in explaining choices after the age of consent: for example, a young adult might not attend college both because his achievements in secondary school were mediocre (which, according to the view just expressed would be a circumstance) and also because his parents put little value on tertiary education (also a circumstance). Facing these two circumstances, if a low-achieving 18-year-old nevertheless succeeds in going to college, through taking compensatory courses, that would be ascribed to exceptional effort, ceteris paribus.

In both the hours of work and education examples, then, we will often not have an accurate measure of effort. It will be measured with error and bias. Broadly speaking, the authors do not pay sufficient attention to these problems and overlook their practical implications. Since effort measurement does not have the same robustness as circumstance measurement, choosing effort as the conditioning variable as in the tranche approach (see for instance, Peragine, 2004; Peragine and Serlenga, 2008) seems risky. True, circumstances may be only partially described, but generally they are not noisy. Since tranche and type approaches seem incompatible (see below), conditioning on type seems a better choice than conditioning on tranches for a measurement error problem.

4.10.1.4 Age and Sex

The issue of availability of information cannot be raised about age and sex. The problem is how to treat these variables. Under the control view, age and sex are circumstances. Under the preference view, because age and sex are important determinants of preference, they will implicitly enter as factors of effort. Because, under this view, preferences should be respected whatever they are unless they are not well-informed, they are put on the responsibility side of the cut.[34] Of course, as Fleurbaey and Schokkaert (2009) pointed out, we are free, once the true impact of age and sex has been identified econometrically, to test whether it matters to put age and sex on one side or on the other (see García-Gómez et al., 2012 for an application). When we are explaining health, it does not come as a surprise to learn that 45% of the explained variance in health comes from these two demographic variables (see Jusot et al., 2013). This is not the thorniest issue in EOp measurement, but the reader should be aware that the extent of inequality of opportunity may depend on whether or not one includes these variables in the responsibility set. For instance, Almås et al. (2011) put age among the responsibility variables, on the ground that our concern should be with inequality of lifetime earnings. Another solution would be to exit the dual world of the model and to admit that there are variables that are neither under the control of the individual nor for which compensation is due. An example is provided in the health sphere where it is admitted, by most, that health policies cannot erase the impact of demographics. (We should not consider males disadvantaged with respect to females if, due to innate biological factors, their life expectancy is shorter.) For earnings achievement, this stance cannot be easily argued, because differences in returns, linked to gender and perhaps age, may be related to discrimination, which would obviously be a violation of EOp.

As in other domains of econometrics, there is a large issue of what to do with poor data. The mistake to avoid is pretending that a poor data set is rich. Innovative methods exist to deal with missing variables. An important methodological issue that has been raised and partially solved is to deduce what can be said about inequality of opportunity when we know that the observables are far from recovering the process through which the objective has been attained. We should adapt our empirical strategy to the richness of the informational structure of the database. Basically, we can contrast situations from the richest informational setting to the poorest one. In the first situation, we have a good description of the world, that is, a quite comprehensive set of circumstances and some candidates for effort variables. In the second situation, no effort variables are available and individuals can be ranked in broad type categories. We will contrast the methods accordingly.

[34] Of course, if age determines both the outcome directly and indirectly through preferences, it is ad hoc to allocate the impact of age entirely to either circumstances or effort.

4.10.2 The Estimation Phase

4.10.2.1 The Case of a Rich Data Set

The first choice is to decide between parametric and nonparametric estimation. Because, by assumption, there are many observable variables, a parametric estimation will fit the data better (see, Pistolesi, 2009 for a semiparametric estimation). Bourguignon et al. (2007) took the lead regarding the econometric strategy in this case. We should estimate a system of simultaneous equations. The first equation will describe the process of attainment of the outcome. In the income context, it can be called a return equation, the coefficient of each determinant giving the marginal return (in a linear model) of each determinant whether it is a circumstance, effort, or demographic variable. The other equations (one for every effort variable) will relate the effort variable to circumstances and other control variables. In the control view of responsibility variables, we should understand how variables that are outside the control of the individual influence her effort variables. In these "reaction equations" circumstances must be introduced, including market conditions (prices, any market disequilibrium such as the local rate of unemployment for job decisions), and demographics. One supposes that the reaction of individuals to their environments (market and background conditions) may vary across individuals. We should let the coefficients vary according to demographics. The difference in the value of these coefficients, if any, would be interpreted in a different way according to the control versus the preference view. According to the latter, they are preference shifters, whereas according to the former they are driven by circumstances, and belong to the nonresponsibility side of the cut.

We introduce some notation. Let y_i be the outcome of individual i (the original outcome variable or some function of it), C_i the vector of circumstances, $E_i = (e_{i1}, \ldots, e_{ij}, \ldots, e_{ik})$ the vector of effort of dimension k, D_i the vector of demographics, M_i the market conditions prevailing for i, ε_i, the mean-zero residual of the return equation, and o_{ij} the mean-zero residual of the reaction equation of effort j. The other letters employed are for coefficients of both regressions. In the simplest linear model the following equations have to be estimated:

$$y_i = \mu_{y1} + \alpha_c C_i + \alpha_d D_i + \alpha_e E_{i,} + \varepsilon_i, \tag{4.16}$$

$$e_{ij} = \mu_{e_j} + \beta_c C_i + \beta_d D_i + \beta_m M_{i,} + \gamma_{cd} C_i D_i + \gamma_{cm} M_i D_i + o_{ij}, \text{ for each effort variable } j = 1, \ldots, k \tag{4.17}$$

Equation (4.16) is written in a compact way: Coefficients β describe the average reaction of adjusting effort to external conditions, whereas coefficients γ are the "preference shifters" which allow individuals to adjust in a different way according to their age and sex group.

It is plausible that market conditions do not always explain the outcome (for instance, the price of fruit and vegetables may impact the diet, while having no impact on mortality rate). If this is the case, we may have exclusion restrictions that will be helpful to identify the system.

The omitted variables (perhaps IQ or any measure of innate talent) may impact the residuals of all equations. The structure of residuals may follow some common pattern that can be captured by a correlation between disturbance terms. (See table 1 in García-Gómez et al., 2012 for an implementation for mortality outcome.) If the correlation is significant, it may reveal an omitted covariate that matters for the estimation of the full system. However, we cannot tell if the revealed omitted variables are on the circumstances or effort side.

Many authors (e.g., Bourguignon et al., 2007; Trannoy et al., 2010) have argued that the estimation of the full system is not necessary if we are only interested in determining the full impact of circumstances. Estimating the reduced form (4.18) suffices if we want to measure the impact of observable circumstances:

$$y_i = \mu_{y_3} + \delta_c C_i + \delta_d D_i + v_i. \tag{4.18}$$

This statement, however, requires some qualification. Neglecting the shift parameter, it is true that in a linear model $\delta_c = \alpha_c + \alpha_e \beta_c$, due to the Frisch-Waugh theorem, α_c captures the direct effect of circumstances and $\alpha_e \beta_c$ captures the indirect effect of circumstances through effort. (The same goes for demographics.) However, the relation is lost for a nonlinear model, such as a logit or probit specification, even if Jusot et al. (2013) found that the difference between δ_c and $\alpha_c + \alpha_e \beta_c$ is quite small. More importantly, the reduced form (4.18), which has been repeatedly estimated in empirical studies, does not allow the effect of circumstances on outcomes to be mediated by demographics. The information provided by the preference shifters γ introduced in the reaction equations (4.17) is lost. It will be split into the reduced coefficient of circumstances, the reduced coefficient of demographics and perhaps the residual. A solution would be to introduce a cross effect of circumstances and demographics in the reduced equation but, to some extent, the effect of demographics as shifters of preferences will go beyond the cross effect in the structural model. The basic message here is that, with a reduced form, we cannot isolate the effect of demographics as circumstances from the effect of demographics as shifters of preferences, and therefore responsibility variables: to do so, we would need to estimate the full structural model. We recall the claim of Fleurbaey and Schokkaert (2009) that failing to estimate a structural model is costly in terms of the limitations that are thereby imposed in the measurement phase.

We now comment on the impact of omitted variables on the estimation. The coefficients will be biased and cannot be interpreted as causal. An example from health is the presence of lead in a child's home, which could entail health problems for both children and parents. If this variable is missing in the data set, a correlation between the health status of children and parents will be observed, whereas there is no causal link. It would then be unwise to base policy recommendations on the estimates of the structural model

(4.16) and (4.17) or the reduced model (4.18). Other empirical strategies have to be implemented if we want to use the estimates in this way. Regarding the reduced form, it must be clear that the estimate $\hat{\delta}_c$[35] conveys the impact of any unobserved variable correlated with observable circumstances. If these variables are circumstances, this is fine from a correlation viewpoint. We can claim that $\hat{\delta}_c C_i$ gives a fair account of the contribution of all factors linked to observable circumstances to the income of individual i.

The interpretation becomes trickier if all the unobservables correlated with circumstances are not interpreted as circumstances. Let us take the example of innate talent and suppose that an accurate measure is IQ. We have advocated treating IQ, measured before the age of consent, as a circumstance. However, as is clear from surveys and questionnaires (see Section 4.8), opinions are quite diverse on this question. If we follow the self-ownership view, it should be a responsibility variable (i.e., persons would deserve to benefit from their high IQs). Ferreira and Gignoux (2011) have argued that the reduced form will lead (through the computation of $\hat{\delta}_c C_i$) to a lower bound estimate of circumstances. If the missing variables in the reduced form are classified as efforts and are positively correlated to observable circumstances such as IQ, it is the other way round. Instead of having a downward bias, the impact of circumstances would be biased upward. The remedy is not trivial because any other simple solution fails to solve the problem. Estimating a reduced form with only observable effort would convey the impact of circumstances correlated with effort, which conflicts with the message of EOp. Now the estimates given by the structural model will be even more at odds with the ethics of EOp. The impact of unobservable IQ will be split into the various coefficients estimated in the return equation (4.16) plus the residual, meaning that some part of innate talent would be assimilated with responsibility characteristics and some part would be nonresponsibility characteristics. At this stage, we should recognize that since innate talent is a form of luck, the parametric estimation is too restricted to cope with luck (see below).

One of the virtues of the structural model is that it enables one to decompose the impact of the circumstances into a direct and an indirect term (through effort). Bourguignon et al. (2007) and Ferreira and Gignoux (2011) acknowledge that subdecompositions into direct or indirect effects, or into the effects of individual circumstances, would be strongly affected by the presence of omitted variables. Bourguignon et al. (2013) show that it is no so much the magnitude of inequality of opportunity, but rather its decomposition between direct and indirect effects, that will be affected by biased estimates of coefficients of circumstances in both the return and the reaction equations.

We conclude with the interpretation of the residuals of the various equations. We first emphasize that they are not orthogonal to the regressors with omitted variables, which is worrying. That said, the residuals of the reaction equation are close in spirit to the

[35] A circumflexed variable denotes an estimate.

Roemerian effort. They are effort sterilized of the impact of circumstances and external conditions. This leads Jusot et al. (2013) to estimate an equation where we substitute Roemerian effort for effort in equation (4.16), namely:

$$y_i = \mu_{y_4} + \delta_c C_i + \delta_d D_i + \alpha_e O_i + \tau_i, \tag{4.19}$$

where O denotes the vector of residuals of equations (4.17). Due to the Frisch-Waugh theorem, the coefficient of Roemerian effort will be the same as the coefficient of true effort, whereas the coefficients of circumstances and demographics will be augmented by their indirect influence through effort and then equal to the coefficients estimated in the reduced equation (4.18).[36] This enables these authors to offer a decomposition of the inequality into responsibility, nonresponsibility, and demographic parts, in the spirit of Roemer. They contrast the results with the estimates obtained with equation (4.16) where the impact of circumstances is only direct and thus follows Brian Barry's recommendation (individuals should be rewarded for their absolute, not relative, effort).

It should be clear from the previous discussion that the residual of the return equation (4.16) is a mixed bag of error terms and omitted variables, which may be circumstances, effort, or luck variables. Generally, the error term represents a large part of the variance, more than 70% in Björklund et al. (2012) for the residual of the reduced form (4.18). It is quite normal that the explained part remains small on cross-sectional estimation: 30% is already an achievement. Should we assign the residual to the effort or circumstance side? Several views clash here. Roemer and his coauthors over the years put the residual of the reduced equation on the effort side, while Devooght (2008) and Almås et al. (2010) put the residual of the structural return equation on the circumstance side.[37] Lefranc et al. (2009) and Jusot et al. (2013) argue that these solutions are ad hoc. They prefer to maintain the position that we cannot tell what the residual represents. Furthermore, when it represents 50% of the variance or more, putting it on one side or the other will determine the relative magnitude of inequality of opportunity. Consequently, they prefer to discard it in any decomposition analysis and move on with the explained part of the outcome, from (4.16):

$$\hat{y}_i = \hat{\mu}_{y1} + \hat{\alpha}_c C_i + \hat{\alpha}_d D_i + \hat{\alpha}_e E_i \tag{4.20}$$

Parametric methods try to estimate the conditional expectation $\boldsymbol{E}(y|C,E)$.[38] Nonparametric methods are more ambitious because they try to estimate the conditional distribution $F(y|C,E)$. O'Neill et al. (2000) were the first to use a kernel density estimator to

[36] In fact, this is not quite correct if market conditions and shift parameters are introduced as in (4.17). The statement is valid for a simple form of (4.17).

[37] They also present robustness results where the residual belongs to the responsibility set. Almås (2008) considers both alternatives.

[38] **E** denotes the expectation operator.

estimate the distribution of income conditional on parental income. It is not by accident that the authors chose a continuous variable (parental income) to perform a nonparametric analysis. The parametric estimation already offers some flexibility for discrete variables. Pistolesi (2009) borrows a semiparametric estimation technique from Donald et al. (2000). In a nutshell, since the hazard rate is defined as

$$H(y) = \frac{f(y)}{1-F(y)} = \frac{f(y)}{S(y|C,E)},$$

with $S(.|.)$ the conditional survivor function, one can write:

$$f(y|C,E) = H(y|C,E)(S(y|C,E)).$$

The trick is then to estimate a hazard-function-based estimator and introduce covariates using a proportional-hazards model. In a second step, the necessary transformations using the above equation are made to obtain an estimate of the associated conditional density function. It is known that the estimation of duration models is more flexible than of linear models. In substance, Pistolesi estimates the conditional distributions corresponding to Equations (4.16) and (4.17) with this estimation technique.

4.10.2.2 The Case of a Poor Data Set

The distinctive feature of a poor data set is that no effort variable is available, but we may still have a rich set of circumstances and a large sample. We can construct types but we cannot a priori build tranches. The approach here comes from Roemer (1993, 1996, 1998) with his identification axiom. It is the only assumption that enables us to say something about inequality of opportunity in the poor-information case. It is nonparametric in essence, since effort is deduced from the distribution of outcome for a type, $F(y|C)$. Two individuals located at the same quantile of their type-conditional distribution are defined as having exerted the same effort, which will be denoted e_{RO}. Formally, starting from the income-generating process given by

$$y = g(C, E),$$

the Roemer identification axiom (RIA) reads:

$$F_y(g(C,E)|\,C) = F_y(g(C',E')|\,C') \Rightarrow e_{RO} = e'_{RO}$$

By construction, this effort is distributed uniformly over [0, 1] for all types. This way of identifying effort has been used by O'Neill et al. (2000) in a nonparametric setting to depict the opportunity set of an heir defined as the income range that she can reach for all levels of Roemerian efforts belonging to [0, 1]. The opportunity sets are contrasted according to the level of advantage given by the decile of parental income.

This way of identifying effort has also been used by Peragine (2004) to build a tranche approach to EOp where the multivariate distribution is described by a matrix whose typical element is the income for a given type and percentile of the type-conditional income distribution. However, this approach is not immune to the omitted variable problem that was discussed above. As was rightly pointed out by Ramos and Van de gaer (2012), omitted circumstances induce wrong identification of the Roemerian effort unless the unobserved circumstances, after conditioning on observed circumstances, no longer affect income (see their Proposition 6). This is a strong condition that will be rarely be satisfied in empirical work.

The identification axiom may be questionable from an analytical point of view (see Fleurbaey, 1998), because it is not clear how multidimensional effort can be aggregated into one indicator, and luck factors can interact with effort in a complex way. The view that the *distribution* of effort specific to a type is a circumstance makes sense in the control view but not in the preference view. Let us coin this axiom as the *type-independent effort distribution*: the relevant normative effort distribution should be independent of type. This axiom is clearly weaker than Roemer's identification axiom. It has inspired fruitful empirical strategies, both in a parametric and nonparametric setting. In the former case, Björklund et al. (2012) estimated a reduced form as in (4.18) with v_i a Gaussian white noise. They assimilate the distribution of the residual to the distribution of effort. However, the distribution of the residual can vary across types and this variation is a nonresponsibility characteristic. They have corrected for variation in the second moment by adding and subtracting to the regression equation a residual term that has the overall variance. Hence, the relevant effort in each type is renormalized to have the same variance.

In a nonparametric setting, Lefranc et al. (2009) retain this independence view of effort, which is postulated in the Roemer identification axiom, without assuming that we can identify effort with the quantile of the type-conditional income distribution. Let the distribution of effort conditional on type (supposed to be unidimensional) be given by $G(e|C)$. The authors follow Roemer's proposition (see Section 4.3) according to which the accountable effort π is given by the quantile within the effort distribution of an individual's type:

$$\pi = G(e|C). \tag{4.21}$$

Equipped with this conception of effort, they are able to link what we can check (in a poor setting) with what we would want to check if we had all the information about effort. What we can check is obviously the equality of the distribution of income conditional on the observables, here, only the vector of circumstances:

$$\text{For any } (C, C'),\quad F(\cdot|C) = F(\cdot|C') \quad \text{(conditional-distribution equality)} \tag{4.22}$$

We have already stated (see Section 4.5) that we would like luck to be even-handed in a world where all circumstances and effort are observed.

$$\text{For any } (C, C', e)\quad F(\cdot|C,e) = F(\cdot|C',e) = K(\cdot|e) \quad \text{(equal-luck opportunity)} \tag{4.23}$$

This allows the distribution of episodic luck to depend on effort but not on circumstances. Their main result, mathematically obvious but of practical importance, is that a necessary condition for equal-luck opportunity to be satisfied is conditional-distribution equality, if we use relative effort. Mathematically, if we replace e by e_r, in (4.23), then (4.23) implies (4.22). Lefranc et al. (2009) prove that this is still the case if some circumstances are not observed. Checking the conditional-distribution equality on the set of observed circumstances is still necessary for the global EOp condition to be satisfied. These results pave the way for using stochastic-dominance tools[39] to measure the unfairness of the distribution, which we discuss below.

4.10.3 The Measurement Phase

Once a model has been estimated, the question of how to proceed to use the estimates obtained in the econometric phase remains open. Various choices have been proposed concerning three issues: the types versus tranches approach, the direct unfairness (DU) versus the fairness gap (FG), and the inequality index. We will deal with these three approaches in turn.

4.10.3.1 Types Versus Tranches

A way to organize the information in a discrete setting is to construct a matrix in which rows are types and columns effort. An element m_{ij} of the matrix is the outcome for type i and effort level j:

$$\begin{array}{cc} & \textit{Effort } j \\ \text{Type } i & \begin{pmatrix} \square & \cdots & \square \\ \vdots & m_{ij} & \vdots \\ \square & \cdots & \square \end{pmatrix} \end{array}$$

It is important to emphasize that this way of proceeding is correct if and only if the knowledge of circumstances and effort is sufficient to determine the outcome level. It means that, with respect to the decomposition of the process allowed by the regression, the residual is assigned to either effort or circumstances, unless the outcome is replaced by the predicted outcome. In this setting, two principles of compensation can be stated. First, we define a *tranche* as the set of individuals who expend the same degree of effort.

The *tranche-compensation principle* states that the closer each column is to a constant vector, the better. If for some effort (column), the inequality of outcome across types is reduced, and everything else remains unchanged, EOp has been improved.

The *type-compensation principle* states that it is good to transfer from an advantaged type to a disadvantaged type, provided that the ranking of types is respected. Suppose that

[39] It is possible to go beyond stochastic dominance to define the relative advantage of a type (see Herrero et al., 2012, for a proposal involving an eigenvalue of a matrix).

between two types, one is unambiguously better off than the other, that is, the outcomes can be ranked unambiguously according to first-order stochastic dominance. Then a transfer from the dominant type to the dominated type for some effort level, ceteris paribus, is EOp enhancing. This principle can be extended further to a second-order stochastic-dominance test (Lefranc et al., 2009). Indeed if two types have the same average outcome but the first one has a larger variance, any risk-averse decision maker would prefer to belong to the second type and consequently one cannot declare that the two types have the same opportunities in terms of risk prospects. The need to take into account the risk dimension echoes the treatment of heteroscedasticity of the residuals in the parametric case by Björklund et al. (2012). This extension leads to a weak criterion of EOp, which corresponds to a situation of absence of second-order stochastic dominance across types.[40]

Fleurbaey and Peragine (2013) show by the means of an example that the two principles clash. There is no complete ordering of the full domain of (positive) matrices, which respects both principles. If we connect this to the results obtained by Lefranc et al. (2009), it is as if we said that *equal-luck opportunity* conflicts with *conditional-distribution equality*.[41] They claim that a choice should be made between the two principles. Logically this is correct. Empirically, it seems to us, that the conflict is not that deep because the principles are useful in different informational contexts. Either one trusts the information about effort and the tranche-compensation principle is appropriate, or one lacks the information about effort, or believes it is insufficiently reliable because of the omitted-variable problem, and then the type-compensation principle remains available.

Fleurbaey and Peragine (2013) also point out that the tranche-compensation principle clashes with two principles of reward, the principle of natural reward and the principle of utilitarian reward. Ramos and Van de gaer (2012) showed that this incompatibility extends to another principle of reward inspired by a criticism of Roemer against the principle of natural reward. The *principle of inequality adverse reward* requires that a within-type Pigou–Dalton transfer be socially desirable.[42] It seems to us that this kind of conflict should not be overemphasized if we agree to prioritize the principles. If we annihilate the inequality due to circumstances according to the tranche-compensation principle, then in each column, each element is equal to its tranche average before the redistribution took place. Hence, this redistribution according to the tranche-compensation principle respects a simple *natural arithmetic average reward* principle: The arithmetic average income

[40] These two principles have been dubbed *ex ante* (type) and *ex post* (tranche) approaches by Fleurbaey and Peragine (2013). The terms are misleading because *ex post* and *ex ante* usually refer to a situation with uncertainty which is not explicit here.

[41] The comparison is not artificial because to some extent, both principles can be viewed as a ranking adaptation of (4.22) and (4.23).

[42] This principle is incompatible with the tranche-compensation principle but not the type-compensation principle.

difference due to differences in effort should remain invariant to redistribution. At this stage, this principle of reward reduces to the principle of natural reward and no more redistribution is required to comply with the requirements of EOp.

We conclude with an insight borrowed from Ramos and Van de gaer (2012), who remark that if we retain the Roemerian effort, annihilating inequality within the columns of the matrix implies equalizing the prospects for each type, since by construction the distribution of Roemerian effort is the same for every type.

4.10.3.2 DU Versus FG

Almost the same idea appears in the papers of Fleurbaey and Schokkaert (2009) and Pistolesi (2009) concerning how to measure inequality due to circumstances. We will here retain the nomenclature of the former authors, while we are closer to the latter in terms of the definitions. These authors propose two approaches.

DU is computed as the inequality of the counterfactual distribution when one has removed the effect of effort variables, either by suppressing them, or by imputing to each individual a reference value of effort such as the average value. Following are some examples of possible computations of DU, where I denotes some inequality index.

For the reduced form (4.18), a natural choice for DU is to compute the inequality of the conditional expectation of outcomes across types (a solution first proposed by Van de gaer, 1993). Since the regression decomposes the conditional expectation, we get

$$I(\mathbf{E}(y|C_i, D_i)) = I\left(\hat{\mu}_{y3} + \hat{\delta}_c C_i + \hat{\delta}_d D_i\right) \tag{4.24}$$

which is a neat solution chosen by Ferreira and Gignoux (2011). The residual is set to 0, its mean value.

For the more structural model (4.16) or (4.19), where an estimation of the impact of the effort variable has been obtained, it is possible to set the effort variable to 0 or to consider some reference value such as the average effort. The inequality of the conditional expectation of outcome for an average effort level is given by

$$I(\mathbf{E}(y|C_i, D_i, \overline{E})) = I\left(\hat{\mu}_{y1} + \hat{\alpha}_c C_i + \hat{\alpha}_d D_i + \hat{\alpha}_e \overline{E}_i\right), \tag{4.25}$$

where an overbar on a variable denotes a mean. A potential problem for both the above calculations is that the distribution of estimated residuals across types may be type dependent. If so, then the difference in the mean of estimated residuals across types should be taken into account.

The FG measures the gap between the inequality of the actual distribution and the inequality of a counterfactual distribution in which all the effects of circumstantial variables have been removed, either by suppressing them, or by imputing to each individual a reference value of circumstances such as the average one. We give some examples below. If we had estimated a reduced form with only effort variables (something that has not been done in the literature so far), we could have the analog of formula (4.24) with an estimation of the inequality of the expected outcomes across tranches when circumstances are in the residual and have been removed. Computing directly from the data the average outcome of those sharing the same effort, as done by Checchi and Peragine (2010), is a nonparametric way of doing this. The FG is then given by[43]

$$I(y) - I(\mathbf{E}(y|E_i)). \tag{4.26}$$

For the more structural model (4.16) or (4.19), where both effort and circumstances variables are introduced as regressors, we can do better and estimate the FG for a counterfactual distribution where the set of circumstances has been set to a reference value, for example, the average one. Then, one obtains for the FG

$$I(y) - I\left(\mathbf{E}\left(y|\overline{C}_i, \overline{D}_i, E_i\right) = I(y) - I\left(\hat{\mu}_{y1} + \hat{\alpha}_c\overline{C}_i + \hat{\alpha}_d\overline{D}_i + \hat{\alpha}_e E_i\right)\right). \tag{4.27}$$

Bourguignon et al. (2007) propose a similar measure. The problem is, again, how to assign the residual. According to (4.27), the residual has been removed and is considered as measuring a circumstance. The above authors implicitly consider the residual as measuring effort. Another solution is to replace the overall inequality by the explained inequality, that is, remembering that $\hat{y}_i$ is the explained outcome (see Equation (4.20)), to compute:

$$I(\hat{y}_i) - I\left(\hat{\mu}_{y1} + \hat{\alpha}_c\overline{C}_i + \hat{\alpha}_d\overline{D}_i + \hat{\alpha}_e E_i\right), \tag{4.28}$$

a solution chosen by Jusot et al. (2013).

The reference values in (4.26) and (4.27) are somewhat arbitrary and we can compute the formula for different values and then take the arithmetic mean. DU and FG as defined above are defined in absolute value. They can of course be defined in relative terms and be divided by the overall inequality. Several recent empirical studies (e.g., Aaberge et al., 2011; Checchi and Peragine, 2010) perform both estimations of the inequality of opportunity as robustness checks.

The measurement of unjust inequality using DU is linked to the *tranche-compensation principle* as follows: if DU computed according to formula (4.25)[44] for some matrix M is lower than for some other matrix M' for all inequality indices, then M is preferred to M' according to the tranche-compensation principle where the considered transfers are of the Pigou-Dalton sort. Similarly, there is a link between the *type-compensation principle* and the FG. Indeed, if M is preferred to M' according to the type-compensation principle,

[43] Fleurbaey and Schokkaert (2009) are the only who propose to apply the inequality index to the gap. The other authors compute the gap between total inequality and the inequality of the counterfactual distribution.

[44] In a parametric or nonparametric way.

then the FG is lower for M than for M', computed according to (4.27), for all inequality indices when the reference type is different from the two types involved in the Pigou-Dalton transfer. The statement is not as general for FG as for DU since we cannot extend the above statement whatever the reference type, the choice of which is ad hoc. This leads some authors to consider instead a weighted average of the FG. In that case it can be proved that, if M is preferred to M' according to the type-compensation principle, then the weighted[45] sum of the FGs is lower for M than for M', computed according to (4.27), for all inequality indices belonging to the entropy class.[46]

We conclude the discussion of DU and the FG by observing that the concepts in substance are not new as methods of decomposing inequality among its sources. When Shorrocks (1980) advocated the use of the variance, he observed in his conclusion that when one thinks about the contribution of one source to inequality, one can wonder either about how much inequality is left when the impact of this inequality factor is neutralized, or about how much inequality remains when the other sources are equalized. This is exactly the choice available in the literature on EOp measurement. Shorrocks (1980) also observed that when there are two sources (here, the set of circumstances and the set of effort variables) the natural decomposition of the variance given by the covariance of the source with outcome has a nice interpretation: the covariance of a source is just equal to the arithmetic mean of the above two computations. In the context of EOp, this means that the covariance of circumstances with outcome is the arithmetic mean of the DU and FG when the other source is removed in the computations (not put at a reference level). This point was made by Jusot et al. (2013) and by Ferreira and Gignoux (2011) (see their appendix).

4.10.3.3 The Choice of an Index

The entire spectrum of inequality indices has been used by researchers in EOp, perhaps with the exception of Atkinson's indices. One can speculate that the absence of the Atkinson indices is due to EOp's not being a welfarist theory. Lefranc et al. (2008) and Almås et al. (2011) have used the Gini index, and Aaberge et al. (2011) have used the Gini and other rank-dependent measures. Elements of the entropy family have been used by Bourguignon et al. (2007), who picked the Theil index, and Checchi and Peragine (2010), Ferreira and Gignoux (2011), Lefranc et al. (2007, 2012) use the MLD. Pistolesi (2009) and Björklund et al. (2012) are eclectic and use a range of measures. These examples are when the objective is income attainment, and they are relative measures. When the objective is health status (self-assessed health or mortality), it makes

[45] For the statement to be true, the weights cannot be chosen arbitrarily. The weight of a type is given by the weight of this type in the between-type term.

[46] For further results regarding the link between the tranche/type approaches and the DU/FG measures, see Brunori and Peragine (2011).

sense to use an absolute measure such as the variance, a choice made by Jusot et al. (2013) and Bricard et al. (2013), which possesses the decomposition property mentioned above. However, the variance is not such a good choice for income attainment since it is not relative. Returning to the income case, there is no first-best choice. The connection with stochastic dominance, which is the advantage of rank-dependent measures (among them the Gini index), is counterbalanced by the decomposability properties of the entropy family. The relevant decomposition is among sources of inequality, and not so much among subpopulations, and the Shapley decomposition (Chantreuil and Trannoy, 2013; Shorrocks, 2013) can be applied to any inequality index.

The property of path independence of the MLD pointed out by Foster and Shneyerov (2000) has recently been emphasized by Ferreira and Gignoux (2011) to single out this index. Indeed, path independence is interesting in the context of EOp because it can be interpreted as saying that the inequality measured by the DU criterion be equal to the inequality measured by the FG. This proposition has to be qualified. DU is computed as the inequality of the average outcome across types. The FG is obtained by rescaling the distribution of the outcome due to effort by the ratio of average income to average income in a type. This is one among many possibilities for nullifying the impact of circumstantial factors. Thus, if we find this way of neutralizing the impact of circumstantial inequalities appealing for the FG, then we do not have to worry about computing two measures of EOp because they are equivalent (under path independence). We conclude by saying that in the health realm, variance may be a better choice, whereas MLD is prominent for income achievement.

4.11. RESULTS

It is beyond our scope to present a unified treatment of all empirical results. As argued earlier, the estimates of inequality of opportunity are likely a lower bound of the true figure in all cases and the magnitude of the underestimation is inversely related to the richness of the data set. Consequently, the importance of the empirical results has to be gauged by considering the number of types that can be defined with the data set. Intriguing issues that may arouse the curiosity of the readers can be easily identified. First, what is the extent of EOp with respect to overall inequality? What is the contribution of effort to inequality, is it larger than that of circumstances? Is the indirect contribution of circumstances through its impact on effort sizeable? Does it make much difference to follow Roemer's viewpoint in measuring effort, or will using absolute measures of effort give similar results? Among circumstances, what are the most significant? Is there a common pattern among inequalities of opportunity with respect to the objectives of health, education and income? Is there a difference of magnitude in inequality of opportunity between the developed countries and the developing countries? Does the ranking of countries differ when we look at inequality of opportunities versus inequality of outcomes? Do taxes and benefits or other instruments make a large difference when measuring EOp? (i.e., inequality of opportunity for pre-fisc versus post-fisc income.)

Starting from a very coarse definition of types (three levels for father's education, five levels for income), Lefranc et al. (2008) found that Sweden and Norway almost achieve EOp for income, while at the other extreme in the range of Western countries are Italy and the United States, with other European countries in the middle. The qualitative results are similar to those of Roemer et al. (2003). We will take a closer look at the Nordic countries before reporting the results obtained for Italy and the United States. We will then contrast these results with those obtained for Latin America, Africa, and Turkey.

Three thorough empirical studies have studied EOp for income in Scandinavia: Aaberge et al. (2011) and Almås et al. (2011) for Norway, and Björklund et al. (2012) for Sweden. Starting with the latter, the authors claim that they have a fine-grained typology (1152 types), which partitions the sample into types based on parental-income quartile group (four groups), parental education group (three groups), family structure/type (two groups), number of siblings (three groups), IQ quartile (four groups), and body mass index (BMI) quartile at age 18 (four groups).[47] The random sample consists of 35% of Swedish men born between 1955 and 1967 and the outcome is an average of pre-fisc income over 7 years (age group: 32–38). Looking at the graphs of stochastic dominance reveals something that was already present in Lefranc et al. (2008). The income CDFs of the different educational or parental-income types are quite close. The differences are more pronounced for IQ-types. Parametric results reveal that the three most important contributors to inequality of opportunity are parental income, IQ, and the type heterogeneity of the disturbance (which may be due to effort, luck, or unobserved type heterogeneity, because the parental-income and education groups are still large). Looking at the Gini coefficient (the results are a bit sensitive to the measure, as usual), putting IQ aside, the other "social" circumstances account for between 15.3% and 18.7% of the overall Gini. That means that in the counterfactual situation where the only factors of inequality would be these social circumstances, the Gini coefficient would attain a modest value of 0.043 for the oldest cohort. The contribution of IQ represents about 12% of the overall Gini. So far, these results are very impressive and confirm that Sweden is close to reaching a situation of equal opportunity. Still, it remains to be seen if introducing parental income in a continuous way and perhaps education of both mother and father, thus refining the typology, would alter the results significantly.

The results for Norway obtained by Aaberge et al. (2011) are built on a coarser typology (three educational parental levels, to grow up in a large family or not, to be born in a

[47] BMI is measured at a young age. It would be far more controversial to put BMI on the circumstance side for older people. Of course, there are genetic roots of obesity among some subjects, but the main determinant is lifestyle (see the discussion in Bricard et al., 2013).

main city or not, and birth cohort). Tranches are defined by relying upon the Roemer identification axiom. The data come from a rich longitudinal set containing records for every Norwegian from 1967 to 2006, enabling one to build up a permanent income measure. The Gini coefficient of permanent income is as low as 0.17, and the authors graph Pen's parade (the inverses of the permanent income CDFs) for the three educational groups. These inverse CDFs are quite close. The Gini coefficient corresponding to inequality of opportunity is about 0.05 suggesting that opportunity inequality accounts for about 28% of income inequality when the analysis is based on permanent income. Since the typology is coarser than in Björklund et al. (2012) for Sweden, the results so far are compatible with a higher inequality of opportunity and likely a higher contribution of inequality of opportunity to overall inequality. Almås et al. (2010) use a different methodology and the results cannot be easily compared. Nevertheless, we can observe an upper bound for the impact of effort. If we consider the usual candidates for effort variables such as years of education, hours of work (for those who work), working in the public sector, county of residence, and choice of university major, then effort's raw contribution to the Gini in Norway in 1986 is about 25.5% in pretax income when we do not sterilize effort variables of the impact of circumstances. However, the impact of parental background on effort variables is quite small. It represents one Gini point over a Gini of 0.26. It is generally observed that the unexplained part (by circumstances or effort) remains quite large and even dominant in all empirical studies of inequality of opportunity.

Next, we will review results on the "poor achievers" of the EOp class among developed countries, the United States and Italy. Pistolesi (2009) uses panel data, the PSID from 1968 to 2001, and he considers age, race, education of both parents, the region of birth and the occupation of the father as circumstances. The two responsibility variables are the years of education and the hours of work. Their conditional distributions are estimated nonparametrically against the vector of circumstances. Pistolesi then predicts two counterfactual distributions for both educational and working-duration distributions. In the first, the effect of unequal circumstances is removed, whereas each individual is assumed to have exerted the same effort in the second. The circumstances have a weaker impact on hours of work than on education, a finding quite common across empirical studies, and which makes sense. A presentation of the results with the Gini to allow comparisons with previous studies shows that the share of inequality due to circumstances in the DU sense is about 35% for a 5-year average earnings at the mean point of the distribution. It is indisputably higher than in Sweden, but it follows a quite remarkable decreasing trend over the period. If the results were confirmed, it would mean that the increase in inequality that has occurred in the United States is not due to an increase in inequality of opportunity. Checchi and Peragine (2010) study the inequality of opportunity in Italy. There are three circumstances: parents' education (five types), sex, and regions (North, South). What is striking is that with such a coarse typology, they find that inequality of opportunity accounts for about 20% of overall income inequality in Italy—that is, higher than the 16% in Sweden with a much finer typology.

Perhaps the sharpest indication of inequality of opportunity for income and wealth is a high elasticity of the income (or wealth) of fathers and sons.[48] Corak (2013) provides an excellent review of the facts for highly developed countries. The Great Gatsby Curve is a strongly positive relationship between the Gini coefficient of income and intergenerational income elasticity. For a set of OECD countries, the United States, United Kingdom, and Italy have both the highest Gini of disposable household income (about 0.35) and of intergenerational income elasticity (about 0.5); Norway, Finland, and Denmark have the lowest of both measures (about 0.23 for the Gini, and less than 0.2 for the elasticity). According to Corak (2013), the main determinants of the high elasticities are the behavior of mobility at the top and bottom of the distribution. In the United States, more than half of sons of fathers in the top decile are placed, as adults, in the top three deciles; similarly, about one-half of sons of fathers in the bottom decile are placed, as adults, in the bottom three deciles. In the United States, high-income families pour private resources into their children; Corak (2013) reports that these "enrichment expenditures" (books, computers, summer camps, high-quality day care, and private schooling) total about $8,900 per annum per child for families in the top-income quintile, while families in the bottom quintile spend $1,300 per annum per child (2006 figures). An equal-opportunity policy should compensate low-income children with similar resources, publicly financed. We recall that private schools hardly exist in the Nordic countries, which surely contributes to the lower intergenerational income elasticities there.

Next, we will turn to less-developed countries. The Latin American study by Ferreira and Gignoux (2011) provides results that can be compared with previous studies. Circumstances are defined as ethnicity, father's and mother's occupations, and birth region, for Brazil, Ecuador, Guatemala, Panama, Colombia, and Peru. The number of types is more than one hundred for the first four countries and about 50 for the last two countries. The contribution of circumstances to inequality is quite high and it varies quite a lot across the six countries. If we look at income, Guatemala and Brazil have in common a high value of the share explained by observed circumstances, about one-third, followed by Panama (30%) and Ecuador (26%). The contribution of inequality of opportunity to total inequality is about 28% in Peru and only 23% in Colombia. However, these two countries have fewer types, which biases the estimates downward with respect to the other countries. The authors also provide estimates of the contribution of nonresponsibility characteristics to consumption inequality per capita, which may be more similar to permanent income. The degree to which inequality of opportunity explains inequality is even higher for some countries, over 50% for Guatemala. Ferreira et al. (2011) study the

[48] For an empirical study of the correlation between income inequality, inequality of opportunity, and intergenerational mobility at the international level, see Brunori et al. (2013).

case of Turkey, which has roughly the same level of development as Brazil, and find that on a sample of ever-married women aged 30–49, inequality of opportunity accounts for at least 26% of overall inequality in imputed consumption, which is by and large a lower value that those found for Latin American countries, except for Colombia. For African countries we will refer to the study of Cogneau and Mesplé-Somps (2008). The surveys that are selected are the only large sample nationally representative surveys in Africa that provide information on parental background for adult respondents. They cover two countries under Britain's former colonial rule, Ghana and Uganda, and three countries under France's former colonial rule, Ivory Coast, Guinea, and Madagascar. The types are defined by a small number of occupational, educational and geographical circumstances. For the two most developed countries, Ivory Coast and Ghana, the Gini inequality of opportunity index is about 0.15 (the triple of what is found in Sweden) and it represents about one-third of overall inequality (0.45). The information is poorer for other countries but, given the results one has on a comparative basis, one can guess that the share of inequality of opportunity is even higher there.

All in all, it seems that the inequality of opportunity for income is highly correlated with inequality of income. This observation is confirmed by the high correlation (0.67) between these two kinds of inequality, measured by the Gini coefficient for western countries (Lefranc et al., 2008). Moreover, this strong correlation seems a general pattern that does not depend on the outcome chosen. Indeed, working on the Retrospective Survey of SHARELIFE, which focuses on life histories of Europeans aged 50 and over, Bricard et al. (2013) observe a positive correlation of about 0.39 between inequality of opportunity in health and health inequality. Furthermore, since lifestyles are documented in this data set, the authors are able to show that inequalities of opportunity for health status in Europe represent on average half of the health inequalities due to both circumstances and effort (lifestyles). There are, however, large variations across countries. The health indicator in this study is SAH (self-assessed health) but using mortality indicators as in García-Gómez et al. (2012), the importance of lifestyles also comes out as a distinctive feature. These authors use a rich data set for the Netherlands (1998–2007), linking information about mortality, health events, and lifestyles. They estimate a full structural model that reveals strong educational gradients in healthy lifestyles which, in turn, have the expected effect on mortality.

From a dynamic viewpoint, intergenerational mobility is clearly an important measure of EOp. Almost all studies of intergenerational mobility define the classes between which mobility is measured as income classes; see Chapter 9 for a thorough discussion of the literature. We mention only one study here. Lefranc et al. (2007) show that under a loglinear relationship between parent and child earnings, whose slope β is the intergenerational earnings elasticity, and choosing the MLD as inequality index, then the following relation holds:

$$I_t^f = -\alpha_t + \beta_t I_t^p,$$

The MLD among descendants, I_t^f, can be written as an affine function of the mean MLD among the fathers' incomes at date t, I_t^p, which is a circumstance for children. The constant $-\alpha_t$ can be interpreted as residual inequality were there to be no inequality of parental income. We may interpret $\beta_t I_t^p$ as the inequality of opportunity due to the circumstance of parental earnings. Reduction of inequality of opportunity can derive from either a drop in the intergenerational transmission of advantages (β), or from mitigating income inequality in the parental generation. In the case of France, the authors found that the reduction of inequality of opportunity was a consequence only of a decrease of inequality in fathers' incomes without any clear contribution of the intergenerational link.

We are at the very beginning of solid empirical analyses of inequality of opportunity. Analysis has been hampered so far by limitation of data sets and the intricacy of the issue. For each recent paper beginning with Bourguignon et al. (2007), the same ritual sentence appears in the introduction, to the effect that "this set of circumstance and effort variables is richer than those used so far in the existing empirical literature on inequality of opportunity." If this trend continues, we can be optimistic that, in the coming years, data sets will improve, as the stakes become clearer.

4.12. CONCLUSION

The main contribution of the equality-of-opportunity literature to the vast literature on inequality is to point out that the *source* of inequality matters from an ethical viewpoint. Most would agree that effects of circumstances on persons' well-being that are beyond the control of individuals should be rectified, while at least some differential outcomes due to choice are not compensable at the bar of justice. Thus, measures of inequality *as such* are not terribly useful—unless one is a simple outcome-egalitarian, who views all inequality as unjust. To the extent that economists ignore this ethical principle—and popular view—their measurements of inequality will not persuade people to rectify it.

As we said, the theory of equal opportunity involves both an equalizing aspect and a disequalizing one. Some philosophers focus—we believe excessively—on the disequalizing aspect, which induces criticisms of the approach from the left. We mention the work of Scheffler (2003) and Anderson (1999), both of whom criticize what they call "luck egalitarianism" as too focused on individual choice: to this they oppose a view of "democratic equality," which involves treating all persons with equal dignity and respect. Indeed, one would surely be sympathetic to their complaint, if the entirety of the equal-opportunity approach were limited to cases of expensive tastes, whether or not society should pay for the hospitalization of the motor cyclist who crashes having chosen not to wear a helmet, or even with the more socially important issue of the responsibility for smoking-related disease. These examples focus upon the disequalizing aspect of the equal-opportunity view—that the effects of imprudent choices are not compensable in the strict interpretation of the view. However, we believe that the main focus of the EOp view is upon its mandate for *equalization* of outcomes that are due to differential circumstances: Most urgently, at this juncture in history, for eliminating differences in income, health, and educational achievement which are due to the vastly different socioeconomic backgrounds in which children are raised, due in large part to the institutions of our capitalist societies. The bourgeois revolutions, which eliminated feudalism and inequality of opportunity due to arbitrary social status, although not complete (think of caste in India), marked a huge advance in the equalization of opportunities: but they replaced feudal inequality of opportunity with inequality of opportunity due to differential wealth. Of course, ancient forms of inequality of opportunity, due to gender, ethnicity, and race still remain as well. The Nordic social democracies have done most at eliminating inequality of opportunity due to income and wealth.[49]

We have characterized economic development earlier as an elimination of inequality of opportunity due to parental socioeconomic status. Assuming development continues globally, according to this measure, we will eventually replace the most important circumstance with—we conjecture—inequality due to natural talent. Many people in the experiments we reported support the meritocratic view, that returns to natural talent are just. Perhaps, as we succeed gradually in eliminating inequalities of important objectives that are due to differential wealth, the focus will then turn to inequalities due to differential natural talent. This would not necessarily require that untalented people be compensated for not having access to the pleasure which talented people enjoy from exercising their talents, but it may well require that no income advantage accrue to the talented. (The taxman will not bill you because you get great pleasure from singing in the shower.) Think of the communist slogan, "From each according to his ability, to each according to his need." That slogan does not begrudge the psychological pleasure and social respect that talent garners, but advocates a complete separation of *income* from talent.

Skeptics will say that markets will always be necessary in large and complex societies, and markets cannot operate efficiently if earnings are too sharply divorced from productive contribution. But this view accepts without question the assumption that individuals always maximize selfishly against the tax regime, or other redistributive policy, which they face. In other words, the incentive problem, so central to economic theory today, takes that problem as a fact of nature, like Newton's laws of gravitation. It is, however, not a fact of that kind, but rather a corollary to a particular human psychology, that has

[49] One should also query, of those who advocate "democratic equality" over the kind of equality of opportunity discussed here, whether democratic equality of the kind they envisage can possibly exist before the invidious inequalities due to circumstances are eliminated. How can people treat each other as equals when massive material inequalities among them, due to luck, continue to exist?

developed in a particular historical epoch, when material scarcity is still prevalent globally, and capitalist economic relations are virtually ubiquitous.[50] It is quite possible (and we believe it to be so) that human material needs are limited, and an historical period will arrive, perhaps relatively soon, when they are more or less universally satisfied. Keynes (1930) in fact argued that such an epoch was virtually upon us, at least in what he called the progressive countries, and that attitudes toward material acquisition would change radically over the next century. If and when this occurs, it seems to us quite reasonable to conjecture that societies will attempt to eliminate differential rewards to talent, having by then done away with inequalities due to feudal status, and capitalist wealth. The question of how an economic mechanism can accomplish this efficiently may well be the central problem for economists of that era.

ACKNOWLEDGMENTS

We thank Tony Atkinson, François Bourguignon, Marc Fleurbaey, and Erik Schokkaert for their comments on previous drafts of his chapter. Alain Trannoy thanks the support of the A*MIDEX project (n° ANR-11-IDEX-0001-02).

REFERENCES

Aaberge, R., Mogstad, M., Peragine, V., 2011. Measuring long-term inequality of opportunity. J. Public Econ. 95, 193–204.
Alesina, A., Angeletos, G.M., 2005. Fairness and redistribution. Am. Econ. Rev. 95 (4), 960–980.
Almås, I., 2008. Equalizing income versus equalizing opportunity: a comparison of the United States and Germany. Res. Econ. Inequal. 16, 129–156.
Almås, I., Cappelen, A., Sorensen, E., Tungodden, B., 2010. Fairness and the development of inequality acceptance. Science 328 (5982), 1176–1178.
Almås, I., Cappelen, A.W., Lind, J.T., Sørensen, E.R., Tungodden, B., 2011. Measuring unfair (in)equality. J. Public Econ. 95 (7–8), 488–499.
Anderson, E., 1999. What is the point of equality? Ethics 109, 287–337.
Arneson, R., 1989. Equality and equal opportunity for welfare. Philos. Stud. 56, 77–93.
Arneson, R.J., 1990. Liberalism, distributive subjectivism, and equal opportunity for welfare. Philos. Public Aff. 19, 158–194.
Arrow, K., Bowles, S., Durlauf, S. (Eds.), 2000. Meritocracy and Economic Inequality. Princeton University Press, Princeton, NJ.
Atkinson, T., Bourguignon, F., 1987. Income distribution and differences in needs. In: Feiwel, G.F. (Ed.), Arrow and the Foundation of the Theory of Economic Policy. Macmillan, London.
Bargain, O., Decoster, A., Dolls, M., Neumann, D., Peichl, A., Siegloch, S., 2013. Welfare, labor supply and heterogeneous preferences: evidence for Europe and the United States. Soc. Choice Welf. 41, 789–817.
Barry, B., 1991. Theories of Justice. University of California Press, Berkeley.
Björklund, A., Jäntti, M., Roemer, J., 2012. Equality of opportunity and the distribution of long-run income in Sweden. Soc. Choice Welf. 39, 675–696.
Bossert, W., 1995. Redistribution mechanisms based on individual characteristics. Math. Soc. Sci. 29, 1–17.
Bossert, W., 1997. Opportunity sets and individual well-being. Soc. Choice Welf. 14, 97–112.

[50] We do not claim that humans have no propensity to be self-interested, but rather that propensity may be vastly overblown. It is difficult to know how human psychology will change as material scarcity fades into the past.

Bossert, W., Fleurbaey, M., 1996. Redistribution and compensation. Soc. Choice Welf. 13, 343–355.
Bourguignon, F., Ferreira, F.H.G., Menendez, M., 2007. Inequality of opportunity in Brazil. Rev. Income Wealth 53, 585–618.
Bourguignon, F., Ferreira, F.H.G., Menendez, M., 2013. Inequality of opportunity in Brazil: a corrigendum. Rev. Income Wealth 53, 551–555.
Bowles, S., 1973. Understanding economic opportunity. Am. Econ. Rev. 63, 346–356.
Bricard, D., Jusot, F., Trannoy, A., Tubeuf, S., 2013. Inequality of opportunity in health and the principle of natural reward: evidence from European countries. Res. Econ. Inequal. 21, 335–370.
Brunori, P., Peragine, V., 2011. Compensation, reward, and the measurement of unfair inequalities. Res. Econ. Inequal. 19, 1–21.
Brunori, P., Ferreira, F., Peragine, V., 2013. Inequality of Opportunity, Income Inequality and Economic Mobility: Some International Comparisons, IZA DP No. 7155.
Buchanan, J.M., 1986. Liberty, Market and State: Political Economy in the 1980s. New York University Press, New York.
Calsamiglia, C., 2009. Decentralizing equality of opportunity. Int. Econ. Rev. 50, 273–290.
Camerer, C.F., 2003. Behavioral Game Theory: Experiments in Strategic Interaction. Princeton University Press, Princeton.
Cappelen, A., Hole, A., Sorensen, E., Tungodden, B., 2007. The pluralism of fairness ideals: an experimental approach. Am. Econ. Rev. 97, 818–827.
Cappelen, A., Sorensen, E., Tungodden, B., 2010. Responsibility for what? Fairness and individual responsibility. Eur. Econ. Rev. 54, 429–441.
Cappelen, A., Konow, J., Sorensen, E., Tungodden, B., 2013. Just luck: an experimental study of risk taking and fairness. Am. Econ. Rev. 103, 1398–1413.
Chanel, O., Luchini, S., Teschl, M., Trannoy, A., 2013. Experimental Redistributive Justice: Disentangling Fairness Views about Responsibility Cut and Selfishness, WP Amse no. 2013–49.
Chantreuil, F., Trannoy, A., 2013. Inequality decomposition values: a trade-off between marginality and efficiency. J. Econ. Inequal. 11, 83–98.
Checchi, D., Peragine, V., 2010. Inequality of opportunity in Italy. J. Econ. Inequal. 8, 429–450.
Cogneau, D., Mesplé-Somps, S., 2008. Inequality of opportunity for income in five countries of Africa. Res. Econ. Inequal. 16, 99–128.
Cohen, G.A., 1989. On the currency of egalitarian justice. Ethics 99, 906–944.
Cohen, G.A., 1993. Equality of what? On welfare, goods, and capabilities. In: Nussbaum, M., Sen, A. (Eds.), The Quality of Life. Oxford University Press, Oxford.
Cohen, G.A., 1997. Where the action is: on the site of distributive justice. Philos. Public Aff. 26, 3–30.
Cohen, G.A., 2004. Expensive taste rides again. In: Burley, J. (Ed.), Dworkin and His Critics. Blackwell, Oxford.
Cohen, G.A., 2006. Luck and equality: a reply to Hurley. Philos. Phenomenol. Res. 70, 439–446.
Cohen, G.A., 2009. Why Not Socialism? Princeton University Press, Princeton.
Conlisk, J., 1974. Can equalization of opportunity reduce social mobility? Am. Econ. Rev. 64, 80–90.
Corak, M., 2013. Income inequality, equality of opportunity, and intergenerational mobility. J. Econ. Perspect. 27, 79–102.
Devooght, K., 2008. To each the same and to each his own: a proposal to measure responsibility-sensitive income inequality. Economica 75, 280–295.
Dolan, P., Tsuchiya, A., 2009. The social welfare function and individual responsibility: some theoretical issues and empirical evidence. J. Health Econ. 28, 210–220.
Donald, S.G., Green, D.A., Paarsch, H.J., 2000. Differences in wage distributions between Canada and the United States: an application of a flexible estimator of distribution functions in the presence of covariates. Rev. Econ. Stud. 67, 609–633.
Dworkin, R., 1981a. What is equality? Part 1: Equality of welfare. Philos. Public Aff. 10, 185–246.
Dworkin, R., 1981b. What is equality? Part 2: Equality of resources. Philos. Public Aff. 10, 283–345.
Ehrlich, I., Becker, G.S., 1972. Market insurance, self-insurance, and self-protection. J. Polit. Econ. 80, 632–648.
European Union Statistics on International Living Conditions (EU-SILC), 2005. epp.eurostat.ec.europa.eu.
Ferreira, F.H.G., Gignoux, J., 2011. The measurement of inequality of opportunity: theory and an application to Latin America. Rev. Income Wealth 57, 622–657.

Ferreira, F.H.G., Gignoux, J., Aran, M., 2011. Measuring inequality of opportunity with imperfect data: the case of Turkey. J. Econ. Inequal. 9, 651–680.
Fleurbaey, M., 1995. Three solutions for the compensation problem. J. Econ. Theory 65, 505–521.
Fleurbaey, M., 1998. Equality among responsible individuals. In: Laslier, J.-F., Fleurbaey, M., Gravel, N., Trannoy, A. (Eds.), Freedom in Economics: New Perspectives in Normative Economics. Routledge, London, pp. 206–234.
Fleurbaey, M., 2008. Fairness, Responsibility, and Welfare. Oxford University Press, Oxford.
Fleurbaey, M., 2012. Equal opportunity, reward and respect for preferences: reply to Roemer. Econ. Philos. 28, 201–216.
Fleurbaey, M., Maniquet, F., 2006. Fair income tax. Rev. Econ. Stud. 73, 55–83.
Fleurbaey, M., Maniquet, F., 2011. A Theory of Fairness and Social Welfare. Cambridge University Press, Cambridge UK.
Fleurbaey, M., Peragine, V., 2013. Ex ante versus ex post equality of opportunity. Economica 80, 118–130.
Fleurbaey, M., Schokkaert, E., 2009. Unfair inequalities in health and health care. J. Health Econ. 28, 73–90.
Fleurbaey, M., Schokkaert, E., Luchini, S., Muller, C., 2013. Equivalent incomes and the economic evaluation of health care. Health Econ. 22, 711–729.
Foley, D., 1967. Resource allocation and the public sector. Yale Econ. Essays 7, 45–98.
Foster, J., 2011. Freedom, opportunity and well-being. In: Arrow, K., Sen, A., Suzumura, K. (Eds.), Handbook of Social Choice and Welfare, vol. 2. Elsevier, Amsterdam.
Foster, J.E., Shneyerov, A., 2000. Path independent inequality measures. J. Econ. Theory 91, 199–222.
Fried, B.H., 1998. The Progressive Assault on Laissez-Faire: Robert Hale and the First Law and Economics Movement. Harvard University Press, Cambridge MA.
Frohlich, N., Oppenheimer, J., Eavey, C., 1987. Choices of principles of distributive justice in experimental groups. Am. J. Pol. Sci. 31, 606–636.
Frohlich, N., Oppenheimer, J., Kurki, A., 2004. Modeling other-regarding preferences and an experimental test. Public Choice 119, 91–117.
Gaertner, W., Schokkaert, E., 2012. Empirical Social Choice. Cambridge University Press, Cambridge.
Gaertner, W., Schwettmann, L., 2007. Equity, responsibility and the cultural dimension. Economica 74, 627–649.
García-Gómez, P., Schokkaert, E., Van Ourti, T., Bago d'Uva, T., 2012. Inequity in the Face of Death. CORE Working Paper no. 2012/14.
Harsanyi, J., 1955. Cardinal welfare, individualistic ethics and interpersonal comparisons of utility. J. Polit. Econ. 63, 309–321.
Heckman, J., 2011. The American family in black and white: a post-racial strategy for improving skills to promote equality. Daedalus 140, 70–89.
Herrero, C., Mendez, I., Villar, A., 2012. Relative Equality of Opportunity: A Statistical Approach. https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=SAEe2012&paper_id=169.
Herrnstein, R., 1971. IQ. Atlantic Monthly (September), 43–64.
Hurley, S., 2002. Roemer on responsibility. Law Philos. 21, 39–64.
Hurley, S., 2003. Justice, Luck and Knowledge. Harvard University Press, Cambridge MA.
Jensen, A., 1969. How much can we boost IQ and academic achievement? Harvard Ed. Rev. 39, 1–123.
Jusot, F., Tubeuf, S., Trannoy, A., 2013. Circumstances and efforts: how important is their correlation for the measurement of inequality of opportunity in health? Health Econ. 22 (12), 1470–1495.
Kahneman, D., Wakker, P., Sarin, R., 1997. Back to Bentham? Explorations of experienced utility. Q. J. Econ. 112, 375–406.
Keynes, J.M., 1930. Economic possibilities for our grandchildren. In: Keyes, J.M. (Ed.), Essays in Persuasion. W.W. Norton, New York, 1963.
Kolm, S.-C., 1972. Justice et équité. Ed. du CNRS, Paris.
Konow, J., 1996. A positive theory of economic fairness. J. Econ. Behav. Organ. 31, 13–35.
Konow, J., 2000. Fair shares: accountability and cognitive dissonance in allocation decisions. Am. Econ. Rev. 90 (4), 1072–1091.
Konow, J., 2001. Fair and square: the four sides of distributive justice. J. Econ. Behav. Organ. 46, 137–164.
Konow, J., 2003. Which is the fairest one of all? A positive analysis of justice theories. J. Econ. Lit. 41, 1188–1239.

Le Grand, J., 1991. Equity and Choice. Harper Collins Academic, London.
Lefranc, A., Pistolesi, N., Trannoy, A., 2007. Une réduction de l'inégalité des chances dans l'inégalité du revenu salarial en France? Rev. Econ. Polit. 117, 91–117.
Lefranc, A., Pistolesi, N., Trannoy, A., 2008. Inequality of opportunities vs. inequality of outcomes: are western societies all alike? Rev. Income Wealth 54, 513–546.
Lefranc, A., Pistolesi, N., Trannoy, A., 2009. Equality of opportunity and luck: definitions and testable conditions, with an application to income in France. J. Public Econ. 93, 1189–1207.
Lefranc, A., Pistolesi, N., Trannoy, A., 2012. Measuring circumstances: Francs or ranks, does it matter? In: Rodriguez, J.G. (Ed.), Inequality of Opportunity: Theory and Measurement.In: Research on Economic Inequality, 19, pp. 131–156, Chapter 6.
Lu, I., Chanel, O., Luchini, S., Trannoy, A., 2013. Responsibility cut in Education and Income acquisition: An Empirical Investigation, Amse WP no. 2013-47.
Messick, D.M., Sentis, K., 1983. Fairness, preference and fairness biases. In: Messick, D.M., Cook, K.S. (Eds.), Equity Theory: Psychological and Sociological Perspectives. Praeger, New York, pp. 61–94.
Moreno-Ternero, J., Roemer, J., 2008. The veil of ignorance violates priority. Econ. Philos. 24, 233–257.
Nozick, R., 1974. Anarchy, State and Utopia. Basic Books, New York.
O'Neill, D., Sweetman, O., Van de gaer, D., 2000. Equality of opportunity and kernel density estimation. In: Fomby, T.B., Hill, R.C. (Eds.), In: Advances in Econometrics, 14. JAI Press, Stamford, pp. 259–274.
Ooghe, E., Schokkaert, E., Van de gaer, D., 2007. Equality of opportunity vs. equality of opportunity sets. Soc. Choice Welfare 28, 209–230.
Otsuka, M., 2011. On the Currency of Egalitarian Justice and Other Essays in Political Philosophy. Princeton University Press, Princeton.
Paes de Barros, R., Ferreira, F.H.G., Molinas Vega, J.R., Chanduvi, J.S., 2009. Measuring Inequality of Opportunities in Latin America and the Caribbean. World Bank, Washington, DC.
Pazner, E., Schmeidler, D., 1978. Egalitarian equivalent allocations: a new concept of economic equity. Q. J. Econ. 92, 671–687.
Peragine, V., 2004. Ranking income distributions according to equality of opportunity. J. Econ. Inequal. 2, 11–30.
Peragine, V., Serlenga, L., 2008. Higher education and equality of opportunity in Italy. Res. Econ. Inequal. 16, 67–97.
Pignataro, G., 2012. Equality of opportunity: policy and measurement paradigms. J. Econ. Surv. 26, 800–834.
Pistolesi, N., 2009. Inequality of opportunity in the land of opportunities, 1968–2001. J. Econ. Inequal. 7, 411–433.
Rakowski, E., 1993. Equal Justice. Clarendon Press, Oxford.
Ramos, X., Van de gaer, D., 2012. Empirical Approaches to Inequality of Opportunity: Principles, Measures, and Evidence, Working Paper Universiteit Gent, 2012/792.
Rawls, J., 1958. Justice as fairness. Philos. Rev. 67, 164–194.
Rawls, J., 1971. A Theory of Justice. Harvard University Press, Cambridge, MA.
Roemer, J., 1985. Equality of talent. Econ. Philos. 1, 151–188.
Roemer, J., 1993. A pragmatic theory of responsibility for the egalitarian planner. Philos. Public Aff. 22, 146–166.
Roemer, J., 1996. Theories of Distributive Justice. Harvard University Press, Cambridge MA.
Roemer, J., 1998. Equality of Opportunity. Harvard University Press.
Roemer, J., 2004. Equal opportunity and intergenerational mobility: going beyond intergenerational income transition matrices. In: Corak, M. (Ed.), Generational Income Mobility in North America and Europe. Cambridge University Press, New York.
Roemer, J., 2012. On several approaches to equality of opportunity. Econ. Philos. 28, 165–200.
Roemer, J., 2013. Economic development as opportunity equalization. World Bank Econ. Rev. 28, 189–209. http://dx.doi.org/10.1093/wber/lht023.
Roemer, J., Ünveren, B., 2012. Dynamic Equality of Opportunity.
Roemer, J.E., Aaberge, R., Colombino, U., Fritzell, J., Jenkins, S.P., Lefranc, A., Marx, I., Page, M., Pommer, E., Ruiz-Castillo, J., San Segundo, M.J., Tranaes, T., Trannoy, A., Wagner, G., Zubiri, I., 2003. To what extent do fiscal systems equalize opportunities for income acquisition among citizens? J. Public Econ. 87, 539–565.

Rosa-Dias, P., 2009. Inequality of opportunity in health: evidence from the UK cohort study. Health Econ. 18, 1057–1074.
Rosa-Dias, P., 2010. Modelling opportunity in health under partial observability of circumstances. Health Econ. 19, 252–264.
Rosa-Dias, P., Jones, A., 2007. Giving equality of opportunity fair innings. Health Econ. 16, 109–112.
Rutström, E.E., Williams, M.B., 2000. Entitlements and fairness: an experimental study of distributive preferences. J. Econ. Behav. Organ. 43, 75–89.
Scanlon, T., 1986. The significance of choice. In: McMurrin, S.M. (Ed.), The Tanner Lectures on Human Values. University of Utah Press, Salt Lake City.
Scheffler, S., 2003. What is egalitarianism? Philos. Public Aff. 31, 5–39.
Schokkaert, E., 1999. Everyone is post-welfarist. Rev. Econ. 50, 811–831(In French).
Schokkaert, E., Capeau, B., 1991. Interindividual differences in opinions about distributive justice. Kyklos 44, 325–345.
Schokkaert, E., Devooght, K., 2003. Responsibility-sensitive fair compensation in different cultures. Soc. Choice Welf. 21, 207–242.
Schokkaert, E., Lagrou, L., 1983. An empirical approach to distributive justice. J. Public Econ. 21, 33–52.
Schokkaert, E., Overlaet, B., 1989. Moral intuitions and economic models of distributive justice. Soc. Choice Welf. 6, 19–31.
Sen, A., 1980. Equality of what? In: McMurrin, S. (Ed.), The Tanner Lectures on Human Values. University of Utah Press, Salt Lake City.
Shorrocks, A., 1980. The class of additively decomposable inequality measures. Econometrica 48, 613–625.
Shorrocks, A., 2013. Decomposition procedures for distributional analysis: a unified framework based on the Shapley value. J. Econ. Inequal. 11, 99–126, as a Rediscovering Classics.
Temkin, L., 1993. Inequality. Oxford University Press.
Trannoy, A., Sandy, T., Jusot, F., Devaux, M., 2010. Inequality of opportunities in health in France: a first pass. Health Econ. 19, 921–938.
Tversky, A., Kahneman, D., 1991. Loss aversion in riskless choice: a reference dependent model. Q. J. Econ. 106, 1039–1061.
Ubel, P., Baron, J., Asch, D., 1999. Social responsibility, personal responsibility, and prognosis in public judgments about transplant allocation. Bioethics 13, 57–68.
Van de gaer, D., 1993. Equality of Opportunity and Investment in Human Capital. Ph.D. dissertation, Leuven University.
Van de gaer, D., Vandenbossche, J., Figueroa, J.L., 2012. Children's Health Opportunities and Project Evaluation: Mexico's Oportunidades Program, CORE DP 2012/15.
Van Parijs, P., 1997. Real Freedom for All. Oxford University Press, New York.
Varian, H.R., 1975. Distributive justice, welfare economics, and the theory of fairness. Philos. Public Aff. 4, 223–247.
Weymark, J., 1991. A reconsideration of the Harsanyi-Sen debate on utilitarianism. In: Elster, J., Roemer, J. (Eds.), Interpersonal Comparisons of Well-Being. Cambridge University Press, Cambridge.
World Bank, 2005. World Development Report 2006: Equity and Development. World Bank, Washington, DC.
Yaari, M., Bar-Hillel, M., 1984. On dividing justly. Soc. Choice Welf. 1, 1–24.

CHAPTER 5

Polarization

Jean-Yves Duclos*, **André-Marie Taptué**†
*IZA, Département d'économique and CIRPÉE, Université Laval, Quebec, QC, Canada
†Département d'économique and CIRPÉE, Université Laval, Quebec, QC, Canada

Contents

Handbook of Income Distribution, Volume 2A
ISSN 1574-0056, http://dx.doi.org/10.1016/B978-0-444-59428-0.00006-0

Abstract

This chapter reviews the basic conceptual foundations for the measurement of polarization, the origins of those foundations, how polarization is distinct from inequality and other ways of considering distances and differences across individuals, and how polarization can be measured in an economic, a social, and a hybrid socioeconomic perspective. The chapter focuses largely on concepts and measurement, with only cursory overviews both of the empirical polarization literature and of the theoretical polarization/conflict literature. The chapter distinguishes five types of polarization: income polarization (where the polarizing variable of interest is any one-dimensional cardinal variable), income bipolarization (the extent to which a population is polarized across two separate groups lying on either side of an income median), social polarization (for cases in which variables of interest are qualitative or have no particular cardinal content), socioeconomic polarization (where some income groups split along social characteristics), and multidimensional polarization (where identity and distances/alienation are measured by several variables of interest).

Keywords

Polarization, Bipolarization, Inequality, Middle class, Alienation, Identification, Social conflict

JEL Classification Codes

A33, D31, D63, D74

5.1. INTRODUCTION

This chapter reviews the basic conceptual foundations for the measurement of polarization, the origins of those foundations, how polarization is distinct from inequality and other ways of considering distances and differences across individuals, and how polarization can be measured in an economic, a social, and in a hybrid socioeconomic perspective. The chapter focuses largely on concepts and measurement, with only cursory overviews of both the empirical polarization literature and the theoretical polarization/conflict literature.

It is useful to stress at the outset that the term *polarization* means different things to different people. First, some people view polarization as important for ethical reasons and others consider polarization as instrumental toward generating tensions and conflicts. Second, there are different "types" of polarization. This chapter distinguishes five such types: income polarization, income bipolarization, social polarization, socioeconomic polarization, and multidimensional polarization. They are listed in Table 5.1, along with their main distinguishing features—how they form groups and how they measure distances. Some of the relevant indices are also listed.

Table 5.1 A categorization of polarization indices

Types of polarization	Identification	Distance	Indices
Income (clustering of cardinal variable around local means)	Discrete/ continuous	Discrete/ continuous	Esteban and Ray (1994), Duclos et al. (2004), and Esteban et al. (2007)
Bipolarization (clustering of cardinal variable across two groups)	Discrete/ continuous	Discrete/ continuous	Thurow (1984), Levy (1987), Leckie (1988), Foster and Wolfson (2010/1992), Blackburn and Bloom (1995), Apouey (2007), Duclos and Échevin (2005), and Chakravarty and Maharaj (2013)
Social (polarization over noncardinal variables)	Qualitative	0/1	Reynal-Querol (2002), Duclos et al. (2004), and Chakravarty and Maharaj (2011b, 2012)
Socioeconomic (social variables define groups; economic variables yield distances)	Qualitative	Discrete/ continuous	Zhang and Kanbur (2001), Gradín (2000), Duclos et al. (2004), Permanyer (2010), Gigliarano and Mosler (2009), and Permanyer and D'Ambrosio (2013)
Multidimensional (multidimensional generalization of income and socioeconomic polarization)	Discrete/ continuous	Discrete/ continuous	Anderson (2010, 2011) and Gigliarano and Mosler (2009)

5.1.1 Income Polarization

The chapter first discusses income polarization, understood as polarization over the univariate distribution of a cardinal variable of interest. This regards polarization as a clustering of that variable around an arbitrary number of local means. Variables of interest in that context usually measure welfare; income polarization is then polarization over the distribution of a measure of welfare. This measure of welfare is often income, which explains why the term "income polarization" is chosen to denote this class of polarization measures. The variable of interest can also be unrelated to welfare. One can think, for instance, of income polarization over a distribution of political attitudes or over a distribution of geographical locations. As long as these variables have cardinal value, polarization over them will be referred to as income polarization.

The formalization of income polarization has mostly relied in the literature on an identification/alienation framework. Members of the same group identify with each other; members of different groups feel alienation with respect to one another. Income polarization is assumed to be increasing in both aspects: the greater the level of group identification or the greater the level of alienation, the greater the level of polarization.

5.1.2 Bipolarization

An alternative notion of polarization across a cardinal variable of interest is bipolarization. Bipolarization captures distances across two groups. These two groups have usually been defined as lying on either side of a median, thus taken as the middle of a distribution; for this reason, the bipolarization literature is closely linked to the literature on the size of the middle class. But one can also think of bipolarization as being concerned with the distance between two other separate income groups, such as the poor and the nonpoor or the bourgeois and the proletarians. (Those two groups cannot, however, be defined by a variable other than the variable interest; using a variable other than the variable of interest to define groups would generate a measure of socioeconomic polarization—see later discussion.)

Two notions are intrinsic to the nature of bipolarization: the notion of *changes in spreads* from the middle and the notion of *variations in "bipolarity."* An increase in the spreads of incomes from a middle position increases bipolarization. An increase in bipolarity—smaller income distances either among those below or among those above the middle—raises bipolarization. Equivalently, a reduction in the income gaps between any two incomes, both above or below the median, increases bipolarization.

The discussion until now already makes clear that the measurement of polarization generally involves both inequality- and equality-like constituents. The equality-like constituent is the basis of the fundamental conceptual difference between inequality and polarization. Polarization differs from inequality in that the importance of pole (or group) homogeneity carries weight in addition to the importance of heterogeneity across individuals. Increased distances across individuals of different groups increase both inequality and polarization; increased bunching (for income polarization) or increased equality (for bipolarization) across individuals of the same group decreases inequality but raises polarization.

Much of the debate on the differences between inequality and polarization—and on the possible relevance of each in explaining conflict—effectively rests on the nature of the effects of Pigou–Dalton transfers on each of these measures of the income distribution. A regressive Pigou–Dalton transfer increases bipolarization if the transfer takes place across the median; however, such a transfer increases inequality but decreases bipolarization if it occurs entirely on one side of the median.[1] Whether either (or both) of these transfers increases conflict is a matter of debate;

[1] Note that there is also evidence in the welfare economics literature that, when asked about inequality, a majority of respondents in various surveys believe that such same-side-of-the-median regressive transfers should also reduce inequality; see Amiel and Cowell (1992).

the polarization literature generally supports the view that a regressive same-side-of-the-median transfer actually reduces conflict.[2]

5.1.3 Social Polarization

The chapter then turns to social polarization. Social polarization is concerned with polarization over variables that are qualitative or have no particular cardinal content. Social polarization does not use information on distances between individuals or groups; it only takes into account the size of the groups and sets distances between them to a constant. This is not to say that social polarization cannot measure tensions or distances across groups; it does this by focusing on the distribution of group sizes.

What matters for social polarization is not only how many groups there are, but also how salient their sizes are. The social polarization literature argues that, ceteris paribus, the larger the size of another group, the greater the threat felt by a given group (proportionally to the size of the other group). This introduces a fundamental distinction between inequality of 0/1 group membership (also called group fractionalization) and social polarization. Fractionalization increases when two identical groups split into two because there is then greater group membership inequality; social polarization falls following such a change.

5.1.4 Socioeconomic and Multidimensional Polarization

Social polarization bases group identity on social characteristics; it sets distances to a binary 0/1 variable and thus does not make use of cardinal distance information. But a richer analysis of differences across members of different social groups can also sometimes be performed jointly on social and economic indicators. Some income groups may be split along some social characteristics; social groups can exhibit heterogeneity in welfare. The introduction of these joint dimensions leads to two generalizations, socioeconomic polarization and multidimensional polarization.

Socioeconomic polarization makes an asymmetric use of these dimensions. One set of social variables is used for group identification; a second set of economic variables fixes distances.[3] Because of this, the usual properties of the income polarization and income bipolarization settings do not apply. In particular, bipolarization properties of increasing spread and of increasing bipolarity do not hold in a socioeconomic polarization context.

Multidimensional polarization measures can also be designed.[4] Group membership can be based on the entire set of social and economic characteristics; so can the

[2] For an exception to this, see Esteban and Ray (2011a), where an increase in within-group inequality increases conflict: increased inequality raises the likelihood that demonstrators within a group may be compensated for the cost of mobilization by other demonstrators within the same group.

[3] This is reminiscent of issues involved in the measurement of equality of opportunity—see Chapter 4 of this Handbook.

[4] See Chapter 3 of this Handbook for broader techniques for comparing dispersions of multivariate distributions.

measurement of distances across individuals. When this is done, a multidimensional analogue of unidimensional income polarization is obtained. Multidimensional polarization can also be of a socioeconomic type; this is achieved by defining social membership by social characteristics (as is usual) and by measuring distances on the basis of a multivariate distribution of welfare indicators.

5.2. MOTIVATION

As for inequality, part of the motivation for studying polarization is ethical. Unlike inequality, however, the ethical motivation comes from the view that distances and differences across groups—as opposed to across individuals for inequality—are normatively undesirable. The hollowing out of the earnings distribution and the disappearance of the middle class may, for instance, create a more segregated and an intrinsically less good society.

Much of the motivation for the study of polarization also owes to the view that it is closely linked to the "generation of tensions, to the possibilities of articulated rebellion and revolt, and to the existence of social unrest in general" (Esteban and Ray, 1994, p. 820). Linking group formation and the eruption of social unrest has a long history. Aristotle wrote (2350 years ago) that "it is manifest that the best political community is formed by citizens of the middle class, and that those states are likely to be well-administered, in which the middle class is large . . . where the middle class is large, there are least likely to be factions and dissension" (Aristotle-350).

To take just another prominent historical example, the Marxist critique of Hegel's philosophy (see, for instance, O'Malley and Blunden, 1970) has forcefully argued that the emergence of two distinct social classes, the bourgeoisie and the proletariat, leads progressively to class struggles as societies industrialize. More generally, humanity's history is often described as a series of group struggles, those groups being typically fairly well defined according to socioeconomic characteristics, interests, and statuses.

The modern formal conceptualization of economic polarization along such antagonistic lines owes much to Esteban and Ray (1994)—ER. Polarization is defined as the grouping of the population into significantly sized clusters such that each cluster has members with similar attributes, and different clusters have members with dissimilar ones. Views differ, however, on how to measure the importance and the relevance of such grouping and on how and whether, in particular, it is social or economic differences that matter in defining polarization. They also differ as to which of these differences is more likely to generate conflict.[5]

[5] See, for instance, the 2008 special issue of the *Journal of Peace Research* on the linkages between polarization and conflict.

Consider first the case of social polarization. Ethnicity (broadly interpreted to include religious, racial, and linguistic identities) is commonly perceived as an important source of it; ethnic groups are seen as firmly bounded, inclined toward ethnocentrism, hostile to outsiders, and as exhibiting a belief that one's ethnic group is centrally important. This can position ethnicity at the center of politics and development in many societies—it is indeed frequent to observe political parties split along ethnic lines (Horowitz, 1985). A social norm of equality further makes the subjugation of ethnic minorities illegitimate and spurs ethnic groups to compare their standing in society against that of other groups. With various technological and social developments, associated political divisions and resultant conflicts may also be increasingly common (Glaeser and Ward, 2006).

The economic consequences of ethnic identities and of (potential or actual) conflict can be numerous (Montalvo and Reynal-Querol, 2005a). Trust and trade may be restricted to individuals of the same ethnic group; public infrastructure may be ethnically biased; government transfers may disproportionately favor some ethnic groups, and so on.

Underlying the listed social tensions also clearly lies a diversity of economic statuses and interests. It is in fact often argued that sociopolitical markers of tension are just proxies for more fundamental economic determinants of polarization (Creamer, 2007). The evidence in McCarty et al. (2006) suggests, for instance, that the increase in the divergence of economic interests of the major constituencies of the two major American political parties may have led to an increase of polarization in American politics.

Theoretical demonstrations of the role of polarization in politics and governance can take a variety of forms. Rent seeking on the part of the different groups is one of them. Montalvo and Reynal-Querol (2005b) present a simple pure contest game in which agents seek rent by spending resources in favor of a preferred group outcome. The utility distances across the groups are set to a constant, and group sizes are assumed to be equal. The level of resources spent by agents affects the probability of success in the contest; that probability is equal to the share of each group in the total resources allocated to the contest. Because of the interaction between group sizes and group probabilities of contest success, the total resources allocated to the contest are shown to be a function of the social polarization index shown in Equation (5.33).

Polarization can also be modeled to lead to conflict in contexts in which distances across groups matter. A brief formalization of this takes the following form. Individuals in a given group derive greater utility from outcomes preferred by groups that are "closer" to them. The probability of a group implementing its preferred (ideal) outcome at the expense of other groups depends on the resources spent by the group in taking control. The sum of these resources also determines the importance of group conflict.

Outcomes of such a game have been studied in Esteban and Ray (1999). Increases in the utility distance between any pairs of groups lead to increases in social conflict. Conflict is always maximized on the symmetric bimodal distribution of the population. But there are many nonlinearities. For instance, a merger of two groups in a society with at least three social groups can increase or decrease conflict depending on the size of the merged groups as well as on the distribution of the population across the nonmerged groups. A movement away from a symmetric distribution of three equally sized groups to a symmetric two-group distribution initially decreases conflicts before it eventually increases it. Bunching—and not only spreads from a "middle"—emerges as an important determinant of conflicts.

A richer framework for studying both the occurrence and the intensity of conflicts and the effect of polarization and fractionalization in each case is found in Esteban and Ray (2008). In a highly polarized society, conflict is expensive, so that its appearance is rare. But when conflict occurs, it is intense. Less-polarized societies, where conflict is less costly, witness a higher frequency of social unrest but of a more moderate intensity. Frequency and intensity can therefore be negatively correlated; it may be that the overall importance of conflict (frequency times intensity) may be reached at intermediate levels of polarization. The occurrence and the intensity of conflicts also depend crucially on the nature of the political system. For all these reasons, the precise overall relationship between polarization and conflict ends up being complex and nonlinear, even in relatively well-structured theoretical settings.

Group distances and group identity affect differently inequality, social polarization, and income polarization. Group cohesion and group distances can also influence differently the size of group conflict in an environment in which control over resources is partly determined by group action. The most elaborate modeling of the linkages between conflict and features of group distance and identity is found in Esteban and Ray (2011b).

The game consists of fighting over a budget, a fraction of which is used for a public good and the rest of which is used for a private good. Each group has a most-preferred composition of public goods. An exogenous fraction of the budget is used for the public good; the precise allocation of that public good is endogenously determined by the identity of the winning group. A degree of group cohesion captures whether group payoffs, and not individual ones, are maximized. Group cohesion can be interpreted, for instance, as a degree of altruism or as an indicator of group leadership. Both of these features reinforce group cohesion and also provide possible answers to the important question of why individuals would want—or would need—to act in groups.

Group members choose to make contributions to the resources used by their group to increase the group's probability to win the game, seize control over the government's budget, and choose the allocation of the public good. In equilibrium, members of any given group make the same contribution; in equilibrium, the groups might be contributing different per capita levels of effort, but it is shown that this is not expected to affect significantly the value of aggregate conflict intensity. An approximately linear relationship is then established in Esteban and Ray (2011b) between equilibrium conflict intensity, the Gini index, fractionalization, and polarization over differences in public good utilities. Four interesting polar cases emerge.

First, when all goods are private, utility distances across different group choices of public good allocations do not matter. This is because utility distances matter only when the allocation of public goods is important to groups. Only group sizes (and not group distances) drive conflict over the allocation of private goods. Distance-based indicators of divergences such as inequality and polarization do not predict conflict; it is only those indicators that are based solely on group size divergences (namely, fractionalization) that matter.

Second, when all goods are public, distances over preferences become salient, and the distance-influenced measures of polarization and inequality dominate fractionalization (which is invariant to group distances) as determinants of conflict. The degree to which polarization dominates inequality as a determinant of conflict depends on the strength of group cohesion; the greater the degree of group cohesion, the greater the conflict-determining importance of polarization as a measure influenced both by group sizes and by group distances.

Third, the relative importance of polarization and fractionalization depends on the importance of public goods. If public goods dominate the allocation of government spending, then conflict rests entirely on the value of winning the public goods allocation game. The importance of polarization (which takes into account the utility distances across groups) then dominates the importance of fractionalization as a determinant of conflict because it alone takes into account public good utility distances across groups and therefore the value of competing for the precise allocation of that good.

Last, inequality has a negligible role when group cohesion is important and, more important, when population size is large (a large population makes individual action less effective). More generally, conflict arises only if population size is small (in which case inequality across individuals is important because there are few interacting individuals and individual action is thus important) or if there is group cohesion (in which case it is polarization and fractionalization—instead of inequality—that play an important role).

5.3. NOTATION

Before proceeding, it is useful to introduce the common notation that will be used in the different sections of the chapter. We denote income—or any other cardinal measure of welfare or, more generally, any cardinal measure of "locations" along which individuals can be found and that can therefore be used to measure distances across individuals by y. Let $F(y)$ be the cumulative distribution function (*cdf*) of income and $p=F(y)$ the proportion of individuals in the population that enjoy a level of income that is less than or equal to y. We will often suppose the existence of a density function, which is the first-order derivative of a continuous *cdf*, denoted as $f(y)=F'(y)$. For discrete distributions, we will think of $f(y)$ as the relative frequency of an income of a value y.

We will denote quantiles as $Q(p)$. For a strictly increasing continuous $F(y)$, they can be defined as $F(Q(p))=p$, or using the inverse distribution function, as $Q(p)=F^{(-1)}(p)$. For discrete distributions, $Q(p)=\inf\{y : p \le F(y)\}$. Thus, with a discrete distribution of N

incomes y_i, ranked in increasing values of y_i such that $y_1 \leq y_2 \leq y_3 \leq \cdots \leq y_{N-1} \leq y_N$, the quantiles are given by $Q(p) = y_i$, for $(i-1)/N < p \leq i/N$ and for $i = 1, \ldots, N$.

The median $Q(0.5)$ is denoted by m and mean income by μ. The mean-to-median ratio is given by $\widetilde{\mu} = \mu/m$ and is a measure of the skewness of the distribution. Let $\widetilde{Q}(q) = Q(q)/Q(0.5)$ denote quantiles normalized by the median, and let $\widetilde{F}(\cdot)$ be the *cdf* of those median-normalized incomes. The Lorenz curve $L(p)$ for $0 \leq p \leq 1$ gives the share of income received by the poorest p proportion of the population. It is defined by $L(p) = \int_0^p Q(q)\mathrm{d}q/\mu$ for all $p \in [0, 1]$. An example of a Lorenz curve is shown in Figure 5.9, to which we return later for additional details.

We will often need to think of a partitioning of the population into n separate and exclusive socioeconomic groups. We will denote by $\mathbf{F}$ the vector of the nonnormalized *cdf* for these groups, each of them denoted by $F_i(y)$ for group i. By definition, $F(y) = \sum_{i=1}^{n} F_i(y)$. The total population of each of these groups is given by n_i, and the vector of these population sizes is denoted by $\mathbf{n}$. Mean income and the relative population share of group i are given, respectively, by μ_i and $\pi_i = n_i/n$ (these shares sum to 1).

5.4. INCOME POLARIZATION

5.4.1 Discrete Income Polarization

The classic and influential formulation by Esteban and Ray (1994) of an alienation/identification framework also presents the first axiomatic formalization of polarization. Every person has one value of a discrete cardinal (such as a discrete income level) attribute; those with the same value identify with the same group. Different values of the attribute generate alienation between members of different groups. A high degree of homogeneity within each group (also called "internal homogeneity"), a high degree of heterogeneity across groups (external heterogeneity), and a small number of significantly sized groups increase polarization.

More formally, it is assumed that the income distribution can be split into a finite number of income classes $i = 1, \ldots, n$, each with income precisely equal to y_i (another interpretation uses μ_i instead of y_i; see following). Identification felt by individuals is an increasing function $I(n_i)$ of the number n_i of individuals in their income class i. The sense of identification depends only on the number of individuals in a given income cluster. (Extensions of this to account for other characteristics belong to socioeconomic polarization.) The distance of individual i from another individual j is denoted by $\delta(y_i, y_j)$. An individual with income y_i feels alienation $a[\delta(y_i, y_j)]$ toward an individual with income y_j. The effective antagonism felt by i toward j is represented by a continuous function $T(I, a)$, where $a = a[\delta(y, y')]$ and $I = I(n_i)$.

The basic formulation of the Esteban and Ray (1994) polarization index ER is then:

$$\mathrm{ER}(\mathbf{n}, \mathbf{y}) = \sum_{i=1}^{n} \sum_{j=1}^{n} n_i n_j T\{I(n_i), a[\delta(y_i, y_j)]\} \tag{5.1}$$

Hence, polarization within a society depends *only* on the distribution of effective antagonism, $T\{I(n_i), a[\delta(y_i, y_j)]\}$. A justification for the additive formulation of Equation (5.1) may be provided along the lines suggested by Harsanyi (1953): an impartial observer might want to use the expected value of effective antagonism to judge overall polarization.

Esteban and Ray (1994) narrow the preceding general formulation by imposing one condition and three axioms on Equation (5.1). The condition is one of invariance with regard to population size: polarization orderings should not change if population sizes are all multiplied by the same number. This is called condition H (for homotheticity).

Condition H

Consider two discrete income distributions $(\mathbf{n}, \mathbf{y})$ and $(\mathbf{n}', \mathbf{y}')$. If $\mathrm{ER}(\mathbf{n}, \mathbf{y}) > \mathrm{ER}(\mathbf{n}', \mathbf{y}')$, then $\mathrm{ER}(k\mathbf{n}, \mathbf{y}) > \mathrm{ER}(k\mathbf{n}', \mathbf{y}'), \forall k > 0$. This condition leads to a constant elasticity formulation of polarization with respect to population sizes in Equation (5.1). It is a common condition in welfare economics, and it is also commonly found in the polarization literature.

ER's identification/alienation framework then proceeds with three different axioms. The first two axioms deal with the impact of income movements on polarization. These first two axioms help characterize the structure of the dependence of $T(I, a)$ in Equation (5.1) on alienation. The third axiom considers the impact of changes in the size of population groups; that third axiom defines the sensitivity of antagonism to identification.

Axiom ER 1

Consider a population made of three income groups, as shown in Figure 5.1. If the two smaller and closer groups join together at the average of their incomes, polarization increases.

Axiom ER 1 considers the effect of reducing local alienation on total polarization. The average distance between the two groups originally at x and y and the group at 0 does not change. Average alienation does fall, however, because of incomes x and y moving closer. Axiom ER 1 says that the effect of increased identification of the two groups to the

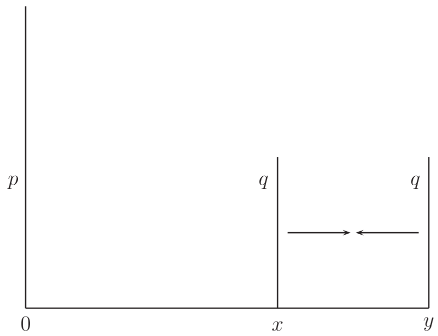


Figure 5.1 Merging two relatively small and close groups at the average of their incomes increases polarization. *Note: There are p individuals at income 0, q individuals at income x, and q individuals at income y. The incomes of the q+q individuals are changed to (x+y)/2.*

right dominates the reduction of alienation. It also implies that the polarization index should be concave in alienation: the effect of the increased alienation between those at incomes 0 and (initial) x dominates the effect of the lower alienation between those at 0 and (initial) y.

Axiom ER 2

Consider a population made of three income groups, as shown in Figure 5.2. If x moves to the right toward y, polarization increases.

Axiom ER 2 involves two changes in alienation. The first one is a fall in alienation between those at x and those at y. The second is an increase in alienation between those at 0 and those at x. Axiom ER 2 says that the greater proximity of x and y should increase polarization. Axiom ER 2 also implies that the index should be convex in alienation. The combinations of Axioms ER 1 and ER 2 imply that the polarization index should be linear in alienation.

Axiom ER 3

Any new distribution formed by shifting population mass from a central mass of size q equally to two lateral masses each of size p and equally distant away from the central mass increases polarization; see Figure 5.3 for an illustration.

Axiom ER 3 says that absorption (or disappearance) of the middle class into richer and poorer classes should increase polarization. It puts a bound on the relative importance of identification in the measurement of polarization: the effect of the fall in identification felt by middle-class individuals located at x in Figure 5.3 should not be too strong in relation to the effect of the increase in identification felt by those individuals on each side of the middle class.

Esteban and Ray (1994) then show that this framework implies a particular form for Equation (5.1):

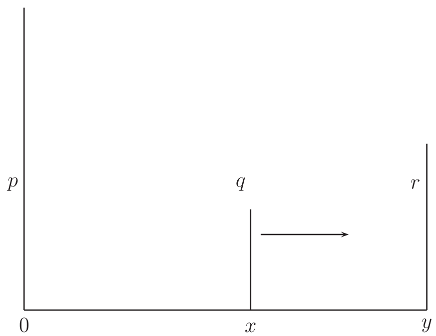


Figure 5.2 Moving x toward y increases polarization. *Note: There are p individuals at income 0, q individuals at income x, and r individuals at income y. Incomes x is moved upward.*

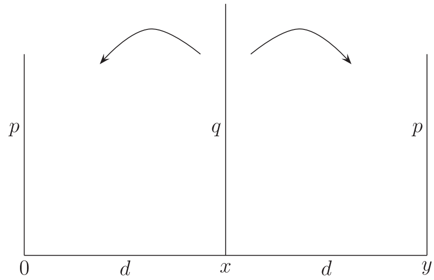


Figure 5.3 Absorption of the middle class into richer and poorer classes increases polarization. *Note: There are p individuals at income 0, q individuals at income x, and p individuals at income y = 2x = 2d. Part of the q individuals at x are shifted equally to 0 and y.*

Theorem 4.1

A class of polarization measure satisfies condition H and the three Axioms ER 1, ER 2, and ER 3 if and only if Equation (5.1) has the form:

$$\mathrm{ER}(\alpha, \mathbf{F}) = K\sum_{i=1}^{n}\sum_{j=1}^{n} n_i^{1+\alpha} n_j |y_i - y_j| \tag{5.2}$$

where $K > 0$ *is a normalization constant and* $\alpha \in (0, 1.6]$, *and if therefore* $T\{I(n_i), a[\delta(y_i, y_j)]\} = n_i^{\alpha}|y_i - y_j|$.

A few remarks may be useful. There are only two degrees of freedom in Equation (5.2), each based on K and α. K is a simple multiplicative constant that has no effect on the ordering of distributions, and α reflects the relative importance of identification and alienation, commonly referred to as a parameter of "polarization aversion." The polarization measure bears resemblance to the Gini coefficient. Aside from the fact that Equation (5.2) uses the logarithm of incomes (see following), it would equal the Gini if α were equal to zero. The fact that α can exceed zero distinguishes income polarization and inequality. The larger the value of α, the greater the departure from inequality measurement because of the greater the departure of Equation (5.2) from the pure consideration of income distances.

In the formulation of Esteban and Ray (1994), y_k in Equation (5.2) is taken to be the log of income (or the log of the relevant cardinal variable). One justification is that individuals may be sensitive to percentage differences in income and not to absolute differences in them. Another justification is that the measure in Equation (5.2) is then invariant to proportional changes in all incomes. The use of income logarithms makes, however, the polarization measure to be noncomparable with the Gini index even when the α parameter is set to zero. Furthermore, $K = \mu^{-1}$ could be used in such a way as to make Equation (5.2) invariant to proportional changes and all incomes, when incomes and not income logarithms are used in the formula.

5.4.2 Continuous Income Polarization

The discrete formula in Equation (5.2) provides an index whose application to indicators of welfare that are commonly continuous (such as income) raises difficulties. First, both the number and the location of income groups are assumed to be (potentially arbitrarily) set/preidentified. Second, a marginal change in the value of y for some individuals may lead to a nonmarginal change in the polarization index (as when a small change in income changes group sizes discretely), a discontinuity that would seem regrettable.

Duclos et al. (2004) (DER) address both of these issues in a framework that is reminiscent of Esteban and Ray (1994) in surface, but with differences at a deeper level. DER postulates that their polarization index should be proportional to the sum of all effective antagonisms in a continuous distribution,

$$\mathrm{DER} = \int\int T[f(x), |x-y|] f(x) f(y) \mathrm{d}x\mathrm{d}y, \tag{5.3}$$

where $f(x)$ is the (unnormalized) density function—capturing identification—and $|x-y| = a$ is the distance between individuals of income x and y—capturing alienation. The antagonism function $T(I, a)$ is increasing in its second argument, and $T(I, 0) = T(0, a) = 0$.

A functional expression for Equation (5.3) is characterized through the formulation of axioms that, though analogous to those of Esteban and Ray (1994), differ in that the income space is continuous. The domains of the axioms are primarily the union of one or more "basic densities." These densities are symmetric, are unimodal, are normalized by population size, and have a compact support. For $0 < r < 1$, a r-squeeze of a density function is defined as a transformation of the form $f^r(x) = (1/r)f[(x-(1-r)\mu)/r]$ such that f^r is a density. The transformation leaves the mean of the distribution unchanged but reduces its standard deviation by a factor $1-r$. The first axiom considers the polarization impact of squeezing basic densities—the first one locally and the second one for two densities, each located on either side of a middle class. These two axioms make it possible to define the role of identification in the antagonism function $T(I, a)$. The third axiom considers the impact of changes in alienation and thus the role of a in $T(I, a)$. The fourth axiom is a population invariance axiom.

Axiom DER 1

A squeeze of a distribution made of only one basic density (as shown in Figure 5.4) does not increase polarization.

Alienation diminishes and identification rises following this squeeze; the impact on polarization of greater identification is nevertheless offset by the impact of a decline in alienation. This effectively limits the parameter α (that will be introduced later, see Equation (5.4), and that is the analogue of the α of the discrete formulation of Equation (5.2)) to be no greater than one because Axiom DER 1 says that polarization should not depart excessively from inequality measurement.

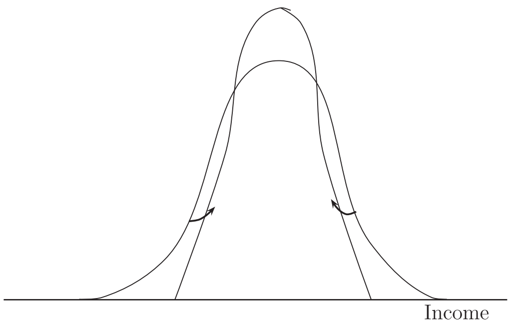


Figure 5.4 A squeeze of a basic density does not increase polarization.

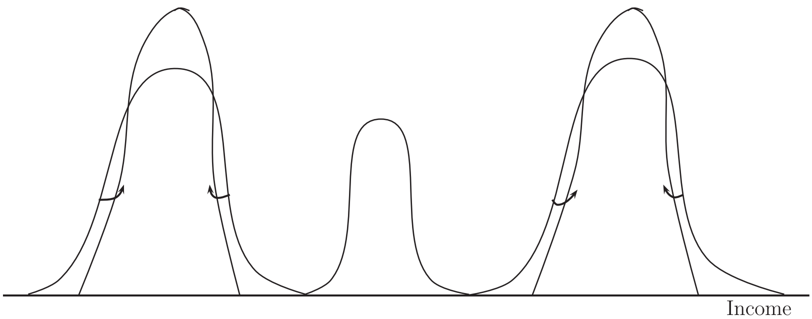


Figure 5.5 A double squeeze cannot reduce polarization.

Axiom DER 2

If the distribution of a symmetric density has three poles (see Figure 5.5), then a squeeze of the outer poles does not decrease polarization.

The decline of intergroup alienation in the outer poles is counterbalanced by the rise of identification. Axiom DER 2 is the "defining" axiom of polarization: It differentiates polarization from inequality. It also puts a lower bound on the parameter α in Equation (5.4): It should not lie below 0.25 for the fall in local alienation (the alienation within each extreme group) to be outweighed by the increase in identification. The axiom also implies that the polarization index should be concave in alienation—for the fall in larger alienation values (the more extreme distances from the middle fall following the squeeze) not to have too large an impact on total polarization.

Axiom DER 3

If a symmetric density has four poles and if each of the two middle poles shifts to the nearer outer pole, as in Figure 5.6, then polarization must go up.

Figure 5.6 shows a fall in local alienation combined with an increase in larger alienation values (an increase in the larger distances between groups). This implies that polarization should be weakly convex is alienation; the effect of a fall in smaller distances should be outweighed by a similarly sized increase in the value of larger distances. Axioms DER 2 and DER 3 imply that polarization should be linear in alienation.

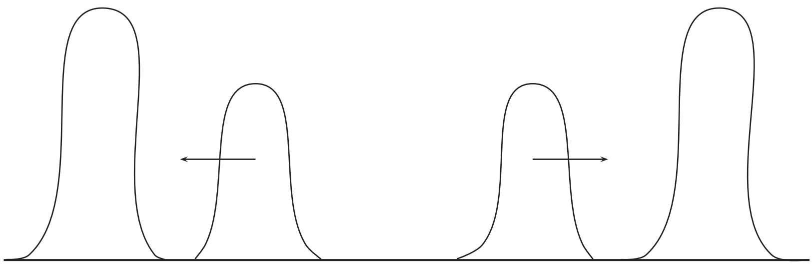


Figure 5.6 A symmetric outward slide must raise polarization.

Axiom DER 4

If the polarization index for one distribution is higher than for another one, then it remains higher when both populations are identically scaled.

This axiom states that polarization orderings should be invariant to population size. It plays a role similar to Condition H given earlier for discrete income polarization. It links identification and polarization through a constant elasticity function in Equation (5.4). DER then shows that:

Theorem 4.2

The index Equation (5.3) satisfies Axioms DER 1, DER 2, DER 3, and DER 4 if and only if it is proportional to:

$$\mathrm{DER}(\alpha) = \int\int f(x)^{1+\alpha} f(y)|x-y|\mathrm{d}x\mathrm{d}y \tag{5.4}$$

for $\alpha \in [0.25\ 1]$, *and if therefore* $T[f(x), |x-y|] = f(x)^{\alpha}|x-y|$.

Making polarization invariant to proportional changes in all incomes can be done by multiplying $\mathrm{DER}(\alpha)$ by $\mu^{\alpha-1}$. For $\alpha=0$, the measure would be equivalent to the Gini coefficient. Note, however, that Theorem 4.2 excludes the Gini index: the Gini index would indeed fall following the squeeze of Figure 5.4, which would violate Axiom DER 2.

The $\mathrm{DER}(\alpha)$ polarization index can be decomposed into identification and alienation components. For a particular value of α, average α-identification can be expressed as

$$\bar{\imath}_{\alpha} \equiv \int f(y)^{\alpha}\mathrm{d}F(y) = \int f(y)^{1+\alpha}\mathrm{d}y. \tag{5.5}$$

The average alienation felt by an individual with income y is given by $a(y) = \int |y-x|\mathrm{d}F(x)$, and the overall average alienation is thus $\bar{a} = \int a(y)\mathrm{d}F(y) = \int\int |y-x|\mathrm{d}F(x)\mathrm{d}F(y)$, which is proportional to the Gini index. Let $\rho_{i,a} = \frac{\mathrm{cov}[f(y)^{\alpha}, a(y)]}{\bar{\imath}_{\alpha}, \bar{a}}$ be the normalized covariance of identification and alienation. Then, $\mathrm{DER}(\alpha) = \bar{\imath}_{\alpha}\bar{a}[1+\rho_{i,a}]$. Ceteris paribus, this means that greater multimodality in the density is likely to translate into greater $\bar{\imath}_\alpha$ and into greater polarization—this effect becoming stronger when α is larger. Greater inequality and thus greater average alienation $\bar{a}$ will also mean greater polarization. Finally, a greater covariance $\rho_{i,a}$ between identification and alienation will raise polarization.

5.4.3 Discrete Income Polarization with Endogenous Grouping

The DER(α) index effectively sets the number of possible income groups to infinity. Each group displays perfect internal homogeneity, being associated with one and only one income level. That removes the need for selecting either the number or the position of income groups.

We might, however, think of measuring income polarization on the basis of a finite and prespecified number of income groups and select their position to maximize internal group homogeneity. Choosing positions to maximize internal homogeneity has the advantage of maximizing local identification and minimizing local alienation—but because individuals in a given group do not all have the same income with discrete income groupings, there will necessarily remain internal heterogeneity, even after such an optimization procedure.

Such an approach is used by Esteban et al. (2007) on the Esteban and Ray (1994) index, using a continuous variable such as income and specifying the number of income groups to be used but not their precise location. Clustering into a finite number of classes introduces errors in the measurement of continuous income polarization. This clustering introduces an approximation error $\varepsilon(\mathbf{F})$. The greater the error, the greater the level of internal heterogeneity, the lower the level of internal identification, the greater the level of local alienation, and the greater the likelihood of a bias in using $\mathrm{ER}(\alpha, \mathbf{F})$ as an index of continuous income polarization. An "extended" polarization index that attempts to correct for such biases is given by:

$$\mathrm{EGR}(\alpha, \sigma, \mathbf{F}) = \mathrm{ER}(\alpha, \mathbf{F}) - \sigma\varepsilon(\mathbf{F}), \tag{5.6}$$

where α is the usual polarization sensitivity parameter and σ is a parameter that weights the measurement error. The approximation error is specified as

$$\varepsilon(\mathbf{F}) = \sum_i \int_x \int_y |x - y| \mathrm{d}F_i(x)\, \mathrm{d}F_i(y) \tag{5.7}$$

and the problem is set to minimize $\varepsilon(\mathbf{F})$ for a given number n of groups. From Equation (5.7), it can be seen that it is equivalent to minimizing the sum of the within-group Gini indices or, alternatively, to minimizing the sum of within-group alienation. The solution is denoted by $\mathbf{F}^*$, with

$$\varepsilon(\mathbf{F}^*) = \mathrm{GB}(\mathbf{F}^*) \tag{5.8}$$

where GB(.) is between-group Gini coefficient of $\mathbf{F}^*$. The Esteban et al. (2007) index is then

$$\mathrm{EGR}(\alpha, \sigma, \mathbf{F}^*) = \mathrm{ER}(\alpha, \mathbf{F}^*) - \sigma \mathrm{GB}(\mathbf{F}^*). \tag{5.9}$$

Note that $\sigma\mathrm{GB}(\mathbf{F}^*)$ is originally meant to capture the effect of one error, the overestimation by $\mathrm{ER}(\alpha, \mathbf{F}^*)$ of income identification in a context in which incomes are artificially grouped. Another error also exists, however, and comes from the underestimation by $\mathrm{ER}(\alpha, \mathbf{F}^*)$ of income alienation in a context in which local alienation is removed through a discrete formation of income groups. The true approximation error of the Esteban and Ray (1994) is therefore quite complicated and should attempt in particular to account both for the identification and the alienation errors of grouping individuals. The formulation of the error should also be consistent with the special form taken by the Esteban and Ray (1994) index; Equation (5.9) takes the awkward shape of a difference between noncomparable functions whenever α is different from zero.

An example of a difficulty created by the form of Equation (5.9) is noted in Lasso de la Vega and Urrutia (2006); the Esteban et al. (2007) measure may fall when groups move farther away from each other because identification errors may increase when this happens, with a resulting fall in Equation (5.9). The alternative index proposed in Lasso de la Vega and Urrutia (2006) is given by

$$\mathrm{CM}(\alpha, \beta, \mathbf{F}) = K \sum_i \sum_j \pi_i \pi_j \pi_i^{\alpha} (1 - G_i)^{\beta} |\ln \mu_i - \ln \mu_j| \tag{5.10}$$

where $\beta \geq 0$ is the degree of sensitivity toward group cohesion, G_i the Gini coefficient of group i, K is a normalization constant, and $1 - G_i$ the Gini *equality* coefficient of group i. The new identification term for each member of group i is thus $\pi_i^{\alpha}(1 - G_i)^{\beta}$, a term that decreases with group Gini. (One could also think of maximizing $\mathrm{CM}(\alpha, \beta, \mathbf{F})$ over a fixed number of groups, in the spirit of the Esteban et al. (2007) index, yielding $\mathrm{CM}(\alpha, \beta, \mathbf{F}^*)$.) As for Equation (5.9), Equation (5.10) does not explicitly account for the dual nature of the identification/alienation approximation error.

5.5. BIPOLARIZATION

5.5.1 Measures of the Size of the Middle Class

Income polarization captures the existence and the importance of an arbitrary number of income poles; bipolarization measures the extent to which a population is divided into two separate groups. An important motivation for developing the concept of bipolarization has been the perception in the 1980s and in the early 1990s—see also Kolm (1969) and Love and Wolfson (1976)—that the size of the "middle class" may be changing over time (and in particular the view that it may be declining). This is because a smaller middle class is presumably associated with greater separateness of the bottom and top halves of the income distribution and with greater (normatively regrettable) distances between groups. Measures of the middle class have, however, relied on definitions that have often been imprecise and heterogeneous.[6]

The size (and the composition) of the middle class is important for a number of economic and social aspects of development. The middle class is a key provider of skilled labor and constitutes an important market for domestic goods and services. The middle class also directly or indirectly provides a large part of a country's tax revenue. A higher share of income for the middle class has been empirically associated with higher incomes and higher growth (Easterly, 2001) as well as with more education, better health care, better infrastructure, better economic policies, less political instability, less civil war and ethnic tension, more social modernization, and more democracy. One reason frequently suggested for this is that a larger-sized middle class is associated with societies with lesser poles at each extreme of the income distribution, thus facilitating political and social harmony and more stable and stronger economic development.

As in the inequality literature, much of the analytical efforts have centered around the construction of indices. The exercise can be made in four different steps: first, specifying the space over which a distribution is split across the middle (population or income wise); second, setting the definition of the middle itself (mean or some middle quantile, such as the median); third, setting the boundaries of the middle class; and, fourth, aggregating the data. An income space is often selected, with income being monthly salary, yearly expenditure, or some other unidimensional indicator of welfare. Though it is common to use the median for the middle, it is also possible to use mean income for that purpose, though the proportion of the population on either side of the mean will typically diverge from 50%, especially if the distribution is not symmetric.[7]

The most influential initial measures of the size of the middle class have then relied on the position of two income cutoff points around the median and on defining the middle class as the share of the population with income within these cutoff points. The middle class is hence defined as the share of the population with incomes between 75% and 125% of the median income by Thurow (1984); Blackburn and Bloom (1995) broaden that middle income range to 60–225% of median income; in Leckie (1988), the middle class is defined on the basis of an income range of 85–115% of median wage. The resulting middle-class indices, denoted as M, is then the share of the population found within these cutoff points. This approach is illustrated graphically using either the density function of

[6] Note that we do not discuss explicitly here the relatively recent literature on the polarization of wages, usually concerned with potentially growing inequality in the top half of the wage distribution and possibly stable or shrinking inequality in the bottom half; see for instance Autor et al. (2008), Acemoglu and Autor (2012), or Mishel et al. (2013).

[7] The middle class can also be defined independent of the "middle." A recent influential example is Kharas (2010), where estimates of the "global" middle class are based on the "ability to lead a comfortable life" and contain those individuals living in households with per capita daily consumption within (PPP) USD 10 and USD 100.

the income distribution (Figure 5.7) or the *cdf* F (Figure 5.8). If the density function is used, the index M is the area situated under its curve and between the lower and the upper bounds of the income range. With the cumulative distributive function, the index M is the vertical distance between the values of the *cdf* evaluated at the two cutoff points.

A people space can also be used. Levy (1987) considers the income share of the middle three-fifths of the population as the size of the middle class. The middle is thus taken to be the 50th percentile, and a range from the 20th to the 80th percentile is identified as the width of the middle class. A middle-class index is then the share of the income earned by that middle class. This can be seen in Figure 5.9 as the difference between two points on the Lorenz curve, $L(0.8) - L(0.2)$. It can also be observed on Figure 5.10 as a quantile function. The measure is the ratio of the hatched area over the total area underneath the quantile curve.

Levy's (1987) index was criticized by Foster and Wolfson (2010/1992) for measuring something other than the size of the middle class or bipolarization. Foster and Wolfson (2010/1992) argued that the index can be considered to be a sound measure of the skewness of the distribution; it should not be viewed, however, as a good index of the size of

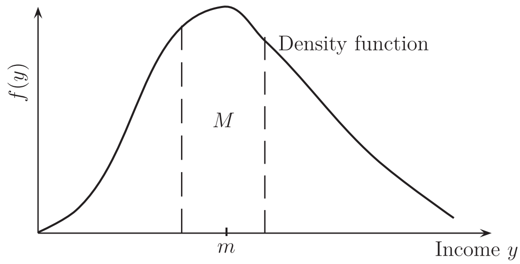


Figure 5.7 Finding the size M of the middle class with a density function.

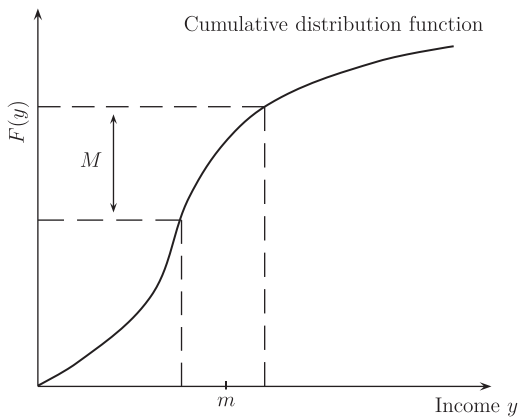


Figure 5.8 Finding the size M of the middle class with a cumulative distribution function.

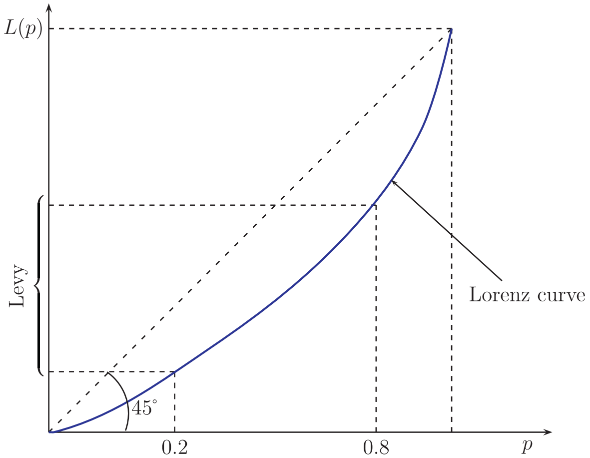


Figure 5.9 Lorenz curve and the Levy (1987) index of the size of the middle class.

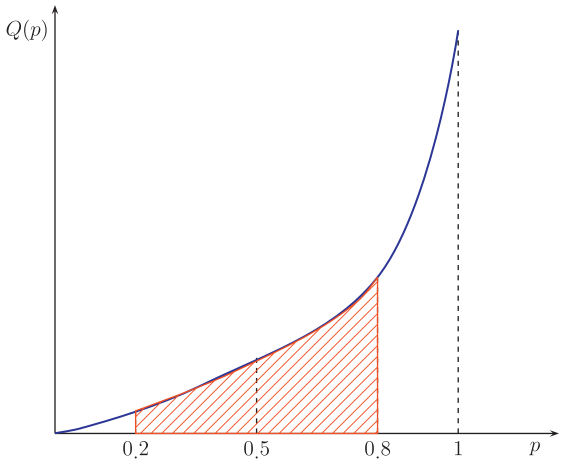


Figure 5.10 Quantile function.

the middle class or of the extent of bipolarization because it fails to measure "spreads" on each side of the middle class. One way to see the problem is to observe that any symmetric distribution will exhibit exactly the same value for the Levy index. For any symmetric distribution, we have indeed that $L(1-p)-L(p)=1-2p$ for all p. For the specific case of $p=0.2$, the index yields a value of 0.6, which is therefore the value of the Levy index for all symmetric distributions, however far incomes may be from the median and however bipolarized these incomes may be.

The fact that Levy-type measures fail to capture spreads on each side of the median can be illustrated in several fashions. For instance, an increase in any quantile

$Q(p)$ for $p \in [0.5, 0.8]$ will increase the Levy measure of the size of the middle class (because $L(0.8)$ will increase and $L(0.2)$ will fall); that, however, will increase the spreads from the median and should therefore conceivably increase bipolarization. Similarly, a fall in $Q(0.1)$ will decrease $L(0.2)$ and increase the Levy measure of the size of the middle class, although that should arguably lead to an increase in the level of bipolarization. Increases in bipolarity that occur entirely within $p \in [0.2, 0.8]$ will not change the Levy measure, although they should reasonably impact bipolarization.

5.5.2 Two Basic Properties of Bipolarization Indices

From the preceding, it is clear that quantifying levels of bipolarization and of the size of the middle class may be treacherous conceptually and thus methodologically. To solve part of the ambiguity, it may be useful to agree on basic properties that the methodological exercise should obey. The two basic properties of bipolarization measurement on which most of the bipolarization literature has insisted are the effects of "increased spread" and "increased bipolarity" (or increased bimodality in Wolfson, 1997).[8]

Consider a discrete population of two income groups, B and D, distributed around median income C and as drawn in Figure 5.11. Now assume that those below and above the median line move away from it; that is, the poorer (at B) become poorer until their income becomes A, and the richer (at D) become richer until their income is at E. Each of these movements away from the middle is called an "increased spread." When this occurs, polarization and inequality are said both to increase. The new distribution is indeed obtained from the old one by a mean preserving regressive transfer across the middle; this increases distances between individuals (and therefore increases inequality), and it

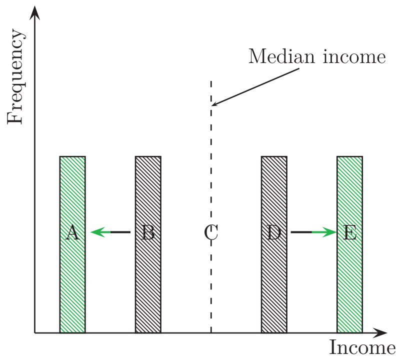


Figure 5.11 Increased spread.

[8] A recent survey of the bipolarization literature can be found in Nissanov et al. (2010).

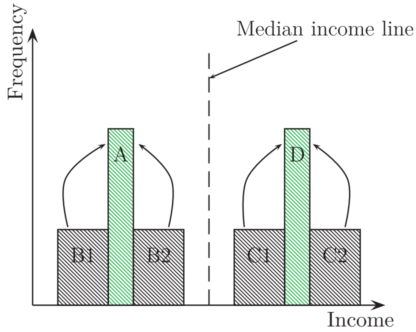


Figure 5.12 Increased bipolarity.

also increases all distances from the middle (and thus raises bipolarization). As in the inequality literature, these movements can be thought to be ethically regrettable.

Figure 5.12 illustrates the property of increased bipolarity. Increased bipolarity is the result of two "spread changes," an "increased spread" and a "decreased spread." Suppose that, below the median, populations with income B1 and B2 cluster at their mean income A and, above the median, populations with income C1 and C2 cluster at their mean income D. Inequality decreases, but polarization can be reasonably said to increase. With increased bipolarity, the average positions of the masses on each side of the median do not change, and the median themselves are not altered, but the distributions on each side of the median are tightened up. During the movement, individuals nearer to the middle move away from it, whereas individuals farther from the median move toward it. The movements of the first group increase the spreads from the median, whereas the movements of the second group of individuals reduce those spreads. Overall, bipolarization increases because the first movements are thought to carry more weight than the second ones.

Increased bipolarity thus distinguishes fundamentally polarization from inequality. Any progressive transfer leads to an unambiguous decrease in inequality, irrespective of the location of the transfer. Bipolarization also falls when the transfer takes place across the middle. Bipolarization is said to increase when the transfer occurs on one side of the middle.[9] A concern about polarization can thus lead us to reject a transfer that would otherwise reduce inequality, where that transfer would take people away from the median.

Much of the literature on bipolarization takes these two properties as given and then proceeds into alternative directions. The first direction leads to the construction of bipolarization indices that are consistent with these two properties. These indices provide complete bipolarization orderings of distributions. The second direction provides dominance curves for making partial bipolarization orderings. The design of these curves draws from the well-known applications of stochastic dominance in the inequality literature.

[9] See Chakravarty (2009, Chap. 4) for a discussion.

It is also on this basis that Foster and Wolfson (2010/1992) (FW) set their influential paper. (The paper was published in 2010 in the *Journal of Economic Inequality* as a "rediscovered classic," having circulated until then as a 1992 working paper.) FW introduce two bipolarization curves as well as a new Gini-like index of bipolarization. The first curve says that bipolarization is greater when the size of the middle class is smaller, viz., when the distances to the median are larger. The second bipolarization curve says that bipolarization is higher when the average distance from the median, on either side of the median, are larger.

The index proposed by FW equals twice the area between the Lorenz curve and the tangent to the Lorenz curve at median income (this area being multiplied by $\widetilde{\mu}$). The index can also be expressed as a function of within- and between-group inequalities: the larger the extent of between-group inequality, the larger is bipolarization, and the larger the level of within-group inequality, the lower is the level of bipolarization. We now examine this in more details.

An income distance between two quantiles at percentiles q^- and q^+ is defined by

$$S(q^-, q^+) = \widetilde{Q}(q^+) - \widetilde{Q}(q^-) \tag{5.11}$$

$S(q^-, q^+)$ is a median-normalized income distance between two quantiles. If $[q^-, q^+]$ overlaps 0.5, it can also be thought of as a measure of the size of the middle class. When S is large, there are fewer individuals near the middle, and consequently the middle class is deemed smaller and there is more bipolarization. FW (see also Wolfson, 1994) define a first-order polarization curve as:

$$S(p) = |S(p, 0.5)| \tag{5.12}$$

for $0 \leq p \leq 1$. For each p, $S(p)$ is the distance that separates median income from the income of the person situated at the pth percentile. This is illustrated in Figure 5.13. Two distances from the median are shown in the upper panel, $S(q^-)$ and $S(q^+)$. The bottom panel of Figure 5.13 mirrors the left-hand side of the upper panel. It draws a first-order bipolarization dominance curve "looking a bit like a lopsided gull" (Wolfson, 1994, p. 355).

FW also define a second-order polarization curve as

$$B(q) = \left| \int_q^{0.5} S(p)\mathrm{d}p \right| \tag{5.13}$$

for $0 \leq q \leq 1$. $B(q)$ is the area under the $S(p)$ curve between points q and 0.5. This is also shown in Figure 5.13 as hatched areas representing $B(q^-)$ and $B(q^+)$. Figure 5.14 shows the shape of a typical $B(p)$ curve, and how its side values $B(0)$ and $B(1)$ can be expressed as functions of the Lorenz point $L(0.5)$. An important point is that $B(q)$ is obtained by integrating "inward" to the median (p from q to 0.5); the usual stochastic dominance curves integrate from 0.

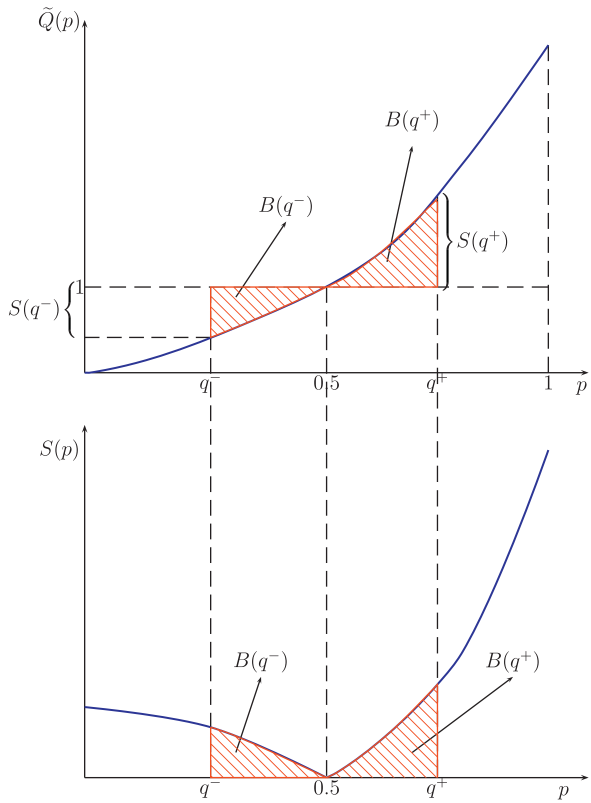


Figure 5.13 First-order bipolarization curves.

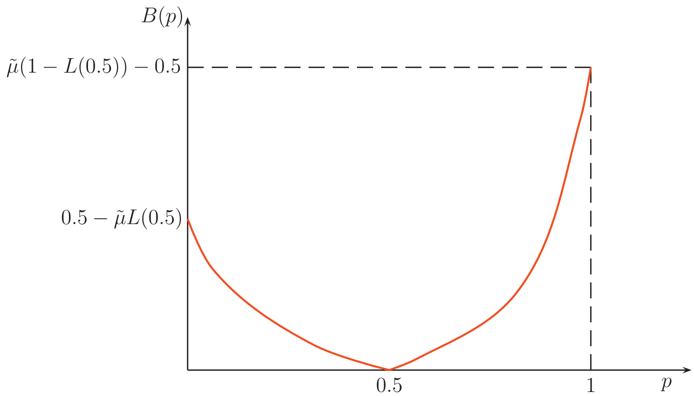


Figure 5.14 Second-order bipolarization curve.

5.5.3 Bipolarization Dominance

The curves just introduced can be used for descriptive as well as for normative purposes. They are called dual curves because they are a function of percentiles p. Analogous dual curves in welfare economics include quantile, Lorenz, generalized Lorenz, poverty gap, and cumulative poverty gap curves—see, for instance, Duclos and Araar (2006) for definitions and a discussion. Primal curves also exist. These primal curves are a function of income levels. These incomes need to be expressed as proportions of the middle (the median, usually) for bipolarization comparisons. A natural first-order primal curve for such comparisons is given by $F(\lambda m) - 0.5$ for $\lambda > 1$, namely, the proportion of the population found between the median itself and λ times the median, or the expression $0.5 - F(\lambda m)$ for $\lambda > 1$, that is, the proportion of the population found between λ times the median and the median. Both of these measures are popular features of the measurement of the size of the middle class—see for instance Morris et al. (1994). They are illustrated in Figure 5.15 (recall that $F(\lambda m) = \widetilde{F}(\lambda)$).

Increases in spreads from the median will decrease each of these two expressions, $F(\lambda m) - 0.5$ and $0.5 - F(\lambda m)$. Increases in bipolarity will increase these expressions at some values of λ and will reduce them at other values of λ.

Let Π^1 then be a class of bipolarization indices P that are functionals of $\widetilde{F}$, population size invariant, monotonically increasing in $\widetilde{Q}(p)$ for $p > 0.5$ and monotonically decreasing in $\widetilde{Q}(p)$ for $p < 0.5$. The following conditions can then be shown to be equivalent:

5.5.3.1 First-Order Bipolarization Dominance

1. $P_B \geq P_A$ for all P in Π^1.
2. $S_A(p) \leq S_B(p)$ for all $0 \leq p \leq 1$.
3. $|\widetilde{Q}_A(p) - 1| \leq |\widetilde{Q}_B(p) - 1|$ for all $0 \leq p \leq 1$.

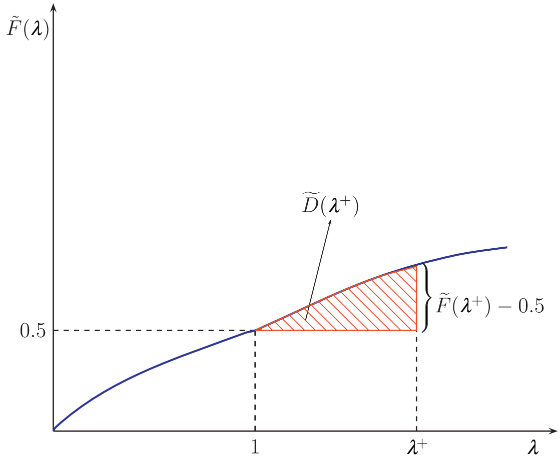


Figure 5.15 Average distance from bipolar extremes.

4. $|\widetilde{F}_A(\lambda) - 0.5| \geq |\widetilde{F}_B(\lambda) - 0.5|$ for all $\lambda > 0$.
5. $|F_A(\lambda m_A) - 0.5| \geq |F_B(\lambda m_B) - 0.5|$ for all $\lambda > 0$.
For $\lambda > 1$, condition 5.4 can be rewritten as $\widetilde{F}_A(\lambda) \leq \widetilde{F}_B(\lambda)$: The proportion of the population with income less than λ times median income (a frequently used relative poverty line) should be lower in A for A to be less bipolarized.

These results can be extended to the case of symmetry of bipolarization indices with respect to distances on each side of the median. Let Π^{S1} be a class of bipolarization indices P that are functionals of $\widetilde{F}(y)$, are population invariant, and are monotonically increasing in $|\widetilde{Q}(p) - 1|$. The following conditions are then equivalent.

5.5.3.2 First-Order Symmetric Bipolarization Dominance

1. $P_B \geq P_A$ for all P in Π^{S1}.
2. $\widetilde{F}_A(\lambda) - \widetilde{F}_A(2-\lambda) \geq \widetilde{F}_B(\lambda) - \widetilde{F}_B(2-\lambda)$ for all $\lambda > 1$.
This is equivalent to comparing the shares of the population within a distance $\lambda - 1$ of the median (Duclos and Échevin, 2005), which is also a frequently used and simple descriptive statistic of the size of the middle class.

The welfare economics literature often stresses the view that some income levels in the income distribution are more important than others, in the sense that changes in these incomes engender greater changes in social evaluation functions (such as social welfare functions or inequality indices). The same view is put forth in the bipolarization literature through the application of the increased bipolarity property, which technically says that indices of bipolarization should be concave in distances from the median and conceptually means that increases in smaller spreads from the median have a greater impact on bipolarization than rises in larger spreads. The curves that are used to establish dominance on the basis of these indices cumulate income distances from the median, in the manner of $B(p)$ for dual curves and in the manner of

$$\widetilde{D}(\lambda) = \begin{cases} \int_{0.5}^{1} \left(\lambda - \widetilde{Q}(p)\right)_{+} \mathrm{d}p & \text{if } \lambda \geq 1 \\ \int_{0}^{0.5} \left(\widetilde{Q}(p) - \lambda\right)_{+} \mathrm{d}p & \text{if } \lambda < 1 \end{cases} \tag{5.14}$$

for primal curves. Figure 5.15 illustrates $\widetilde{D}(\lambda)$ for λ^{+}. It can be seen as the integral of the $F(\lambda m) - 0.5$ (for $\lambda > 1$) and $0.5 - F(\lambda m)$ (for $\lambda < 1$). From Equation (5.14), it can also be seen that $\widetilde{D}(\lambda)$ can be expressed as the sum of the distances between median-normalized incomes $\widetilde{Q}$ and a threshold λ. This is analogous to the aggregation of poverty gaps in the poverty literature: the larger the sum of poverty gaps, the greater is poverty; the larger the sum of the gaps between normalized incomes and a threshold on each side of the median, the greater is bipolarization. $\widetilde{D}(\lambda)$ can also be understood as the average distance from bipolar extremes (λ) on either side of the median.

Let Π^2 be a class of bipolarization indices P that are functionals of $\widetilde{F}$, population invariant, monotonically increasing in $\widetilde{Q}(p)$ for $p > 0.5$, monotonically decreasing in $\widetilde{Q}(p)$ for $p < 0.5$, and concave in $\widetilde{Q}(p)$. The following conditions are then equivalent.

5.5.3.3 Second-Order Bipolarization Dominance

1. $P_B \geq P_A$ for all P in Π^2.
2. $D_A(\lambda) \geq D_B(\lambda)$ for all $\lambda < 0$.
3. $B_A(p) \leq B_B(p)$ for all $0 \leq p \leq 1$.

As discussed earlier, this can be extended to the case of symmetric of bipolarization indices with respect to distances on each side of the median. Let Π^{S2} be a class of bipolarization indices P that are functionals of $\widetilde{F}$, population invariant, and monotonically increasing and concave in $|\widetilde{Q}(p) - 1|$. The following conditions are then equivalent.

5.5.3.4 Second-Order Symmetric Bipolarization Dominance

1. $P_B \geq P_A$ for all P in Π^{S2}.
2. $D_A(\lambda) + D_A(2-\lambda) \geq D_B(\lambda) + D_B(2-\lambda)$ for all $\lambda > 1$.

A natural question is whether these bipolarization partial orderings have power empirically, namely, whether it is possible to rank bipolarization across distributions using them. The comparisons of 29 Luxembourg Income Study (LIS) countries found in Duclos and Échevin (2005) show some evidence of this. The comparisons are made with and without statistical testing of the differences between the curves; it is more difficult to distinguish the curves statistically than numerically.[10] Symmetric and asymmetric bipolarization dominance tests are also performed; symmetric tests have naturally more power. Overall, Duclos and Échevin (2005) found that first-order dominance is found statistically across 32% of the 406 possible pairwise comparisons; that percentage increases to 55% for symmetric first-order dominance testing and to 73% for second-order symmetric dominance.

5.5.4 Bipolarization Indices

Specific numerical indices are frequently used to summarize and compare income distributions and have the advantage of providing complete orderings over distributions. FW propose one such index with two principles in mind:

1. The index should conform to basic underlying notions of the concept being measured; for example, inequality measures should be Lorenz-consistent, the Lorenz curve being the gold standard in comparing inequality robustly and graphically. In the case of bipolarization, FW base their index on the second-order dual bipolarization curve $B(p)$, which is the bipolarization analogue of the Lorenz curve.

[10] See Chapter 6 of this Handbook for a coverage of some of the estimation and inference issues involved in comparing distributions.

2. The index should be easily understandable. For example, the Gini coefficient can be expressed as twice the area between the Lorenz curve and the diagonal of equality. FW follow a similar procedure in proposing their index.

The bipolarization index of Foster and Wolfson (2010/1992) is then defined as

$$\mathrm{FW} = 2\int_0^1 B(p)\mathrm{d}p, \tag{5.15}$$

which is twice the area beneath the second-order bipolarization curve $B(p)$. FW is therefore necessarily consistent with a partial ordering based on comparing $B(q)$. It is also simple to understand.

FW also has quite a few other interesting features. It is, for instance, linked to the Gini index and the line tangent to the Lorenz curve at the median income. Let the average distance between incomes under the median and incomes above the median be given by:

$$T = \left(\mu^{\mathrm{U}} - \mu^{\mathrm{L}}\right)\frac{1}{\mu} = 1 - 2L(0.5), \tag{5.16}$$

where $\mu^{\mathrm{U}} = (1 - L(0.5))\mu$ is the mean income of those above the median and $\mu^{\mathrm{L}} = L(0.5)\mu$ is the mean income of those below the median. FW call T the relative median deviation.[11] $T = 2V$ is twice the share V of total income needed to increase to mean income the incomes of those below the median. Thus, it can be measured as twice the vertical distance from the first diagonal to the Lorenz curve at the level at which the cumulative percentage of the population equals 0.5. This is shown in Figure 5.17.

Another expression for T is given by integrating $S(p)$:

$$T = \widetilde{\mu}^{-1}\int_0^1 S(p)\mathrm{d}p = \widetilde{\mu}^{-1}(B(0) + B(1)), \tag{5.17}$$

so that T is twice the area under the first-order polarization curve normalized by $\widetilde{\mu}$—see Figure 5.13.

T is also twice the area of the quadrilateral 0ABC in Figure 5.17. This has two consequences: first, T is greater than the Gini coefficient G and, second, the area between the Lorenz curve and the tangent line (the light gray area in Figure 5.17) is never zero when G is nonzero.

Foster and Wolfson (2010/1992) then show that

$$\mathrm{FW} = (T - G)\widetilde{\mu} \tag{5.18}$$

The index FW is a scaling up (by the skewness measure (μ/m)) of the light gray area of Figure 5.17. It is thus simple to construct from such basic statistics as mean income, median income, the Gini coefficient, and the relative median deviation.

[11] This is different from the relative mean deviation, which is given by $2(F(\mu) - L(F(\mu)))$.

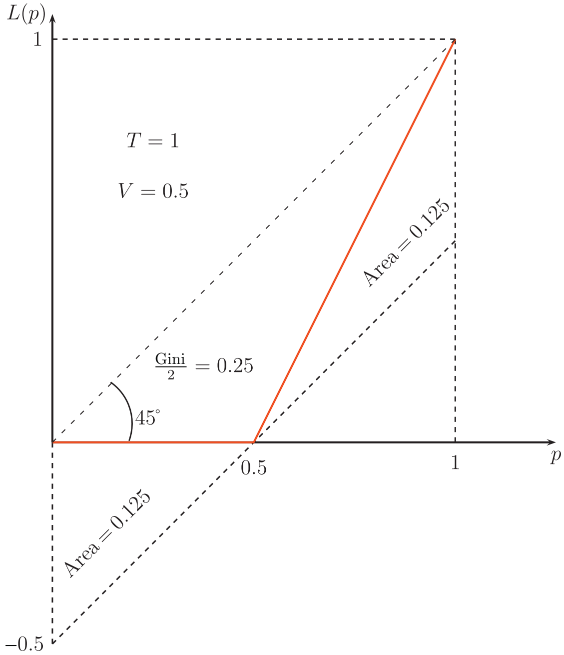


Figure 5.16 Maximum bipolarization.

When a distribution is perfectly bimodal, half of the population has zero income and the other half has income equal to 2μ. This case of perfect bimodality is shown in Figure 5.16. Maximal FW bipolarization is given by twice the area covered by the triangles that lie between the Lorenz curve of the bimodal distribution and the 45° line that touches the horizontal axis at $p=0.5$. Each triangle has an area of 0.125, as shown on the figure. Twice those areas equals 0.5, which is the maximum value attained by FW. FW equals zero in the case of a perfectly equal distribution. Wolfson (1994, 1997) prefer to redefine FW by multiplying it by 2, so that the new index ranges from 0 to 1.

The polarization index FW can also be expressed as a function of components of the Gini coefficient, between-group inequality, and within-group inequality. Divide the population into two groups, one made of individuals with income below the median and the other one made of individuals with income above the median. Give income μ^{L} to individuals below the median and μ^{U} to individuals above the median. GB($\mathbf{F}$) is the between-group Gini coefficient of this new income distribution, which has no within-group inequality. The difference $\mathrm{GW}(\mathbf{F}) = G(\mathbf{F}) - \mathrm{GB}(\mathbf{F})$ is a measure of within-group inequality and is a population-weighted sum of Ginis within the two groups.

An illustration of this decomposition is given in Figure 5.17. The three values of the Lorenz curve at points 0, 0.5, and 1 are joined to form a between-group Lorenz curve, a

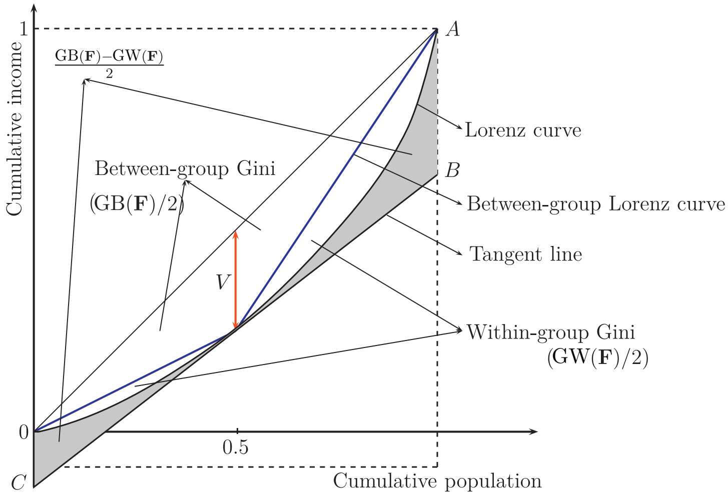


Figure 5.17 Polarization, relative median deviation, and within- and between-group inequality.

piece-wise linear curve. $GW(\mathbf{F})$ is twice the area between the original Lorenz curve and the newly graphed between-group Lorenz curve; $GB(\mathbf{F})$ is twice the area captured by the between-group Lorenz curve, that is, the area between it and the 45° line. The area of the triangle formed by the between-group Lorenz curve and the diagonal equals $[0.5 - L(0.5)]/2$, such that between-group inequality equals $GB(\mathbf{F}) = 0.5 - L(0.5)$ and that $T = 2GB(\mathbf{F})$ is twice the between-group inequality term. Given this, an alternative expression for the FW index is therefore

$$FW = (GB(\mathbf{F}) - GW(\mathbf{F}))\tilde{\mu}. \tag{5.19}$$

Equation (5.19) is a function of between-group inequality minus within-group inequality as measured by the Gini index and with the two groups, respectively, above and below the median. This nicely shows how the bipolarization index FW is influenced both by spreads from the median and by bipolarity. Increases in spreads from the middle raise GW; increases in bipolarity reduce GW. Both effects increase bipolarization. Changes in spreads from the middle have a larger impact on bipolarization when the spreads are initially small: increases in spreads from the middle raise between-group inequality, but within-group inequality indeed falls faster when those closer to the median move away from it.

We use 2004 Canadian data from the LIS to illustrate some of these expressions. The estimates of the Gini for those below and above the median are $G_L = 0.206$ and $G_U = 0.213$, respectively; the data also yield estimates of $\tilde{\mu} = 1.165$ and $L(0.5) = 0.286$. This leads to $T = 0.430$, namely, the relative median deviation and twice the share of total income needed to increase to mean income of those below the median. The between-group Gini index ($GB = 0.5 - L(0.5)$) is found to be 0.215, and the within-group Gini index $(\mathrm{GW} = 0.5L(0.5)G_{\mathrm{L}} + 0.5(1 - L(0.5))G_{\mathrm{U}})$ equals $0.332 - 0.215 = 0.117$. The overall Gini index is therefore $0.215 + 0.117 = 0.332$. By Equation (5.18), this leads to $\mathrm{FW} = 0.114$. The same result is also obtained using Equation (5.19): $\mathrm{FW} = (0.215 - 0.117)\widetilde{\mu} = 0.114$.

As shown in Equation (5.19), inequality and polarization rise together when inequality between the two groups rises; they move in opposite directions when within-group inequality declines. When between-group inequality increases, inequality and polarization rise simultaneously—this corresponds to a greater spread; when within-group inequality ($\mathrm{GW}(\mathbf{F})$) decreases, inequality decreases too, but polarization rises, corresponding to increased bipolarity.

5.5.5 Income Polarization and Bipolarization

Note that increased bipolarity leads to lower inequality on each side of the median, but not necessarily to poles that are necessarily more defined on each of these sides. Consider, for instance, Figure 5.18. The initial distribution (shown by the 45°-hatched rectangles) has four equally sized separate groups, two on each side of the median. Say that the two groups on the right-hand side each split into two smaller groups (leading to the $6 - 45°$-hatched rectangles, two on the left-hand side and four on the right-hand side) following an increase in bipolarity on the right-hand side of the distribution. Within-group inequality has increased, and between-group inequality has been left unchanged. Bipolarization should therefore be judged to have increased, but is this also the case for polarization? It may instead be argued that the poles of the distribution have now become less well defined and that income polarization has fallen.

This important distinction between income polarization and bipolarization can be pushed further. It should be clear that concepts and measurement of income polarization and bipolarization are related and yet different. Conceptually, income polarization is concerned with the existence of multiple groups; bipolarization deals with the existence of

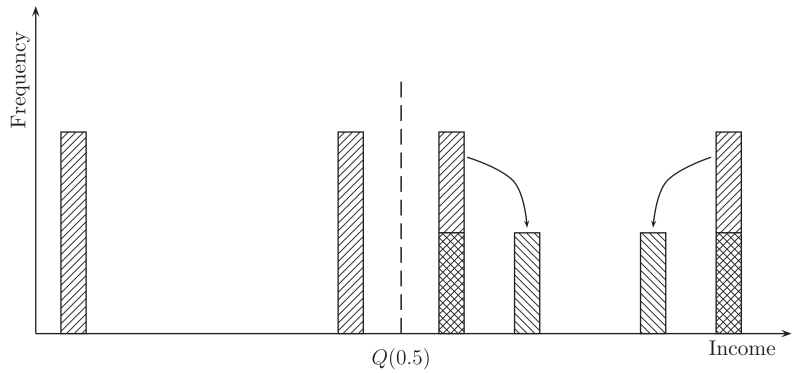


Figure 5.18 Does increasing bipolarity increase polarization?

two bipolar groups. From a measurement perspective, different functional restrictions are also imposed. Income polarization is based on the average antagonism generated by the mixture of identification and alienation. Bipolarization is a function of distances from a middle. It is therefore not surprising that income polarization and bipolarization orderings may clash.

Axiomatically, there are also similarities and differences across the two frameworks. For instance, the properties of bipolarization measurement imply some of the commonly used income polarization axioms. The property of increased spread implies Axioms DER 1 and DER 3; the property of increased bipolarity implies Axiom DER 2. Thus, bipolarization indices such as FW obey Axioms DER 1, DER 2, and DER 3 (and DER 4 as well because of population size invariance). But the converse is not true. The DER axioms do not imply the two fundamental properties of bipolarization measurement. This also suggests that (leaving aside the differences in the initial functional restrictions, see Esteban and Ray, 2012) the income polarization framework is more flexible than the bipolarization one, as intuition also suggests because the income polarization framework is set over an arbitrary number of groupings.

Because of this greater flexibility, the DER indices can fail, however, to obey the increased spread and the increased bipolarity properties (again, see Esteban and Ray, 2012). This is because the movements involved in these properties can decrease identification sufficiently to lead to a fall in average antagonism and thus in income polarization. Again, a demonstration of this is shown in Figure 5.18, where the movements increase bipolarization but may decrease income polarization.

Conversely, bipolarization indices such as FW can fall following squeezes of local distances that increase local identification. For instance, a squeeze of each of three equally sized and equally spaced symmetric basic densities will always decrease FW, but will also always increase DER. (This is in fact true for squeezes of any odd number of such symmetric groups.) Again, the source of the clash comes from the conceptual distinction between distances across well-identified groups and distances of those groups from a middle.

5.5.6 Extensions

Various extensions of the FW index have been proposed. Most of them rely on intuitive alternative applications of the spread/bipolarity framework. One fruitful general approach is to think of distances from the median as the variable of normative interest and to use well-known aggregation techniques from welfare economics to aggregate these distances in a manner that ensures that desirable properties are fulfilled. Note that, for the measurement of polarization, those properties then apply to distances from the median: in welfare economics, the same properties normally apply to distances from the mean or from other incomes—see, for instance, Chapter 4 in Duclos and Araar (2006).

Wang and Tsui (2000) proposed two types of extensions, both of which are axiomatically characterized. In both cases, income distances from the median are aggregated. The first type of aggregation uses rank weighting to aggregate the distances in a manner analogous to Donaldson and Weymark (1980) and Yitzhaki (1983). The measure is defined as

$$\mathrm{WT_A} = \int w(p)S(p)\mathrm{d}p \tag{5.20}$$

where $w(p)$ is positive and increasing for $p \leq 0.5$ and positive and decreasing for $p > 0.5$. Positivity of p is needed for the increasing spread property to hold: Any increase in $S(p)$ must raise $\mathrm{WT_A}$. Increasingness of $w(p)$ for $p \leq 0.5$ and decreasingness of $w(p)$ for $p \geq 0.5$ are needed for the increasing bipolarity property to hold as well. Makdissi and Mussard (2010) used classes of such indices to assess the impact of tax reforms.

Wang and Tsui (2000) also considered polarization measures defined as transformation of distances from the median and normalized in such a way that there is no polarization when income is equally distributed across all individuals. A class of such indices is given by

$$\mathrm{WT_B} = \int_0^1 \psi(S(p))\mathrm{d}p, \tag{5.21}$$

where $\psi(u)$ is a continuous function. The class of polarization indices $\mathrm{WT_B}$ satisfies the axioms of increasing spread and of bipolarity if and only if $\psi(u)$ is strictly increasing, strictly concave, and with $\psi(0)=0$. A constant elasticity formulation for ψ emerges if rankings of nonnormalized distances must always be the same as those for normalized distances from the median. This yields the polarization class of indices

$$\mathrm{WT_C} = \int S(p)^r\mathrm{d}p \tag{5.22}$$

where $r \in (0,1)$. The larger the value of r, the more sensitive are the polarization indices to the deviations of richer persons' incomes from the median.

Inspiration from the inequality literature can also be used to generalize the aggregation of distances from the median that Foster and Wolfson (2010/1992) and much of the subsequent bipolarization literature have adopted. An example of this is Chakravarty and Majumder (2001)'s exploration of the use of the Atkinson (1970), Kolm (1969), and Sen (1973) inequality indices for the measurement of bipolarization—see also Chakravarty (2009, section 4.3). By constructing, inter alia, "equally distributed equivalent incomes" on the separate distributions of incomes lower and greater than the median, it is possible to take into account inequality on each side of the median to assess by how far the two groups are from the middle of the entire distribution. When the welfare evaluation function is of the Gini type, such procedures simplify to the Foster and Wolfson (2010/1992) measure of bipolarization.

Rodríguez and Salas (2003) follow in FW's footpath by applying a sensitivity parameter to weight the subgroup decomposition of the Gini coefficient into between and within-group inequality. Using the Donaldson and Weymark (1980) and Yitzhaki (1983) single parameter/extended Gini coefficients, they propose a bipolarization index of the form

$$\mathrm{RS}(\nu, \mathbf{F}) = \mathrm{GB}(\nu, \mathbf{F}) - \mathrm{GW}(\nu, \mathbf{F}). \tag{5.23}$$

where ν is an inequality aversion parameter—see Duclos and Araar (2006) for a discussion. A progressive median-preserving transfer within (between) the two groups on each side of the median increases (reduces) polarization.

5.5.7 Absolute and Relative Bipolarization Indices

The literature on inequality offers both absolute and relative inequality indices. Absolute inequality indices are invariant to translations of every income by the same constant. Relative inequality indices are invariant to changes in the scaling of all incomes; they are homogeneous of degree zero in all incomes.

Similar distinctions can be (and have been) made in the bipolarization literature. Chakravarty et al. (2007) discuss how the second-order bipolarization curve of Foster and Wolfson (2010/1992) can be made absolute, in the sense of being invariant to translations of every income by the same amount. Integrating the area under this absolute curve leads to an absolute version of the relative bipolarization index of Foster and Wolfson (2010/1992). The absolute inequality indices of the Kolm (1969) and Donaldson and Weymark (1980) types can also be used to weight the sums of absolute distances from the median.

Chakravarty and D'Ambrosio (2010) furthered these distinctions by defining intermediate bipolarization indices, namely, bipolarization indices that generate as special cases absolute and relative bipolarization indices. More precisely, we may want a distribution of income y and a distribution $y + c[\gamma y + (1-\gamma)]$ to exhibit the same level of bipolarization, where $c > 0$ is a scalar. This implies that it is the distances $S(p)m/(\gamma m + (1-\gamma))$ that must be aggregated in the measurement of bipolarization. The bipolarization curves of Equations (5.12) and (5.13) then become intermediate polarization curves and are given by

$$\mathrm{IS}(p, \gamma) = \frac{S(p)}{\gamma + (1-\gamma)/m} \tag{5.24}$$

and

$$\mathrm{IB}(p, \gamma) = \frac{B(p)}{\gamma + (1-\gamma)/m}. \tag{5.25}$$

When $\gamma = 0$, an absolute polarization curve is obtained; $\gamma = 1$ yields a relative polarization curve. These curves can be used to derive an intermediate polarization index analogous to that of Foster and Wolfson (2010/1992), given by the area below the intermediate bipolarization curve. This yields:

$$\mathrm{IFW} = \frac{\mathrm{FW}}{\gamma + (1-\gamma)/m}. \tag{5.26}$$

The intermediate polarization index IFW becomes an absolute index if $\gamma = 0$ and a relative one if $\gamma = 1$.

We may also wish distributional rankings to remain invariant to the choice of measurement units. This invariance property (known as unit consistency) does not require indices to be invariant to changes in monetary units (such as cents or dollars); it only requires the distributional orderings not to be affected by such changes. The implications of such a property for the measurement of bipolarization are explored in Lasso de la Vega et al. (2010), following Zheng (2007) for the measurement of inequality. The results make use of Krtscha-type (Krstcha, 1994) intermediate polarization indices, indices that rank polarization identically across distributions A and B if and only if

$$\frac{Q_{\mathrm{A}}(p) - m_{\mathrm{A}}}{m_{\mathrm{A}}^{\gamma}} = \frac{Q_{\mathrm{B}}(p) - m_{\mathrm{B}}}{m_{\mathrm{B}}^{\gamma}} \quad \text{for all } p. \tag{5.27}$$

$\gamma \in [0,1]$ can be regarded as a degree of bipolarization intermediateness. The extreme values of γ equal to 0 and 1 correspond to absolute and relative polarization measures and orderings. Krtscha-type bipolarization orderings can then be made on the basis of $m^{(1-\gamma)}S(p)$. Lasso de la Vega et al. (2010) show that the only type of bipolarization orderings that are unit consistent are those of the Krtscha type.

5.5.8 Bipolarization with Ordinal Data

It is not difficult to think of applications of bipolarization to contexts in which the welfare variable is discrete and ordinal. Examples of such variables include education, class, and positional statuses as well as health indicators. Although such indicators have ordinal content (values of the variables can be ranked), the fact that they are not cardinal makes it difficult to mean or median normalize them, as is typically done for inequality analysis.

It may, however, still be possible to compare the extent of bipolarization over distributions of such variables. An example of such an exercise is found in Apouey (2007) for distributions of discrete and ordinal self-assessed health (SAH) data.[12] It is first assumed that a distribution exhibits no bipolarization when everybody has the same health condition, and that a distribution exhibits maximal polarization if half the population has the lowest SAH indicator and the other half has the highest one. It is then assumed that bipolarization can be expressed as the sum of the distances $|F(i) - 0.5|$ for each level of category i, $i = 1, \ldots, n$, ranked from lowest to highest values.

[12] See also Chakravarty and Maharaj (2013) for a related setting that uses generalized Gini indices.

It is then shown that

$$\mathrm{Apo}(\eta, F) = K\left[(0.5)^{\eta} - \frac{1}{n-1}\sum_{i=1}^{n-1}|F(i) - 0.5|^{\eta}\right], \tag{5.28}$$

where K is a strictly positive constant and $\eta > 0$ is the weight on the median category, which is the only bipolarization measure that satisfies a population invariance axiom and an increasing spread axiom. Bipolarization is maximal when $F(i) = 0.5$ for all $i < n$. An axiom of increased bipolarity is obeyed if and only if $\eta \in]0, 1[$ and therefore if $\mathrm{Apo}(\eta, F)$ is convex in $F(i)$.

Movements of $F(i)$ away from 0.5 increase bipolarization more if $F(i)$ is initially close to 0.5. An increase in bipolarity moves values of $F(i)$ close to 0.5 away from it and moves more extreme values of $F(i)$ closer to 0.5. Such an increase in bipolarity thus increases bipolarization.

5.6. SOCIAL POLARIZATION

5.6.1 Concepts and Motivation

Social polarization arises when people are clustered according to social characteristics and feel alienated from others that do not share these characteristics. The term is used when the factors that determine individuals' identity are culturally, ideologically, historically, biologically, or socially driven and do not depend solely on their income levels. Religion, race, ethnicity, language, education, and occupation are examples of characteristics along which society may be polarized. The study of social polarization is usually motivated by the view that, in many circumstances, income may not be the only relevant (or even be a relevant) dimension that might nourish identity and differences and thus generate conflict, as suggested by Esteban and Ray (1994), Montalvo and Reynal-Querol (2005b), Collier and Hoeffler (2004), and Easterly and Levine (1997), among many others. This has led to the relatively recent development of several social polarization measures, such as those of Reynal-Querol (2002), Apouey (2007), Permanyer (2010), and Permanyer and D'Ambrosio (2013). These are particularly useful in those contexts in which cardinal data are not available and in which polarization must be built on the basis of qualitative and nonordinal variables.

Religion may be seen as one of the most important of these qualitative characteristics. Reynal-Querol's (2002)—RQ—contention, for instance, is that religiously divided societies are more prone to intense conflicts because religious identity is fixed and nonnegotiable. She analyzed the effect of ethnic division on civil war as well as the role of political systems in preventing these conflicts, using the importance of religious polarization and animist diversity.

Others have focused on ethnic diversity to attempt to explain economic and social phenomena such as growth, investment, government efficiency, civil wars, and unemployment. Easterly and Levine (1997) explained part of the poor economic performance in Africa by the relatively high level of ethnic fragmentation found on the continent. Montalvo and Reynal-Querol (2002) argued that ethnic diversity leads to the choice of poor public policies and a poor design of human/physical capital accumulation in particular, which has a negative influence on long-run growth and encourages rent seeking. Mauro (1995) also contends that ethnic fractionalization increases corruption and political instability and decreases investment. Because corruption has a negative impact on investment, he concluded that ethnic diversity undermines growth, a view shared by Porta et al. (1999).

5.6.2 Measurement

To see how social polarization is measured in the literature, let us first recall the Esteban and Ray (1994) index of Equation (5.2):

$$\mathrm{ER}(\alpha, \mathbf{F}) = K\sum_{i=1}^{n}\sum_{j=1}^{n}\pi_i^{1+\alpha}\pi_j|y_i - y_j|. \tag{5.29}$$

In the case of social polarization, where income plays no role and where each person is identified with every other member of the group, the alienation function assumes values that are specific to group pairs and have no relationship to income. Denoting this alienation value by δ_{jk} for a pair of groups j and k, Duclos et al. (2004) suggested a natural simplification of the ER and DER polarization measures as:

$$\mathrm{DER_S}(\alpha, \mathbf{F}) = \sum_{j=1}^{N}\sum_{k=1}^{N}\pi_j^{1+\alpha}\pi_k\delta_{jk}. \tag{5.30}$$

Substituting the Euclidean distance $\delta(y_i, y_j) = |y_i - y_j|$ by the discrete distance

$$\delta(y_i, y_j) = \begin{cases} 0 & \text{if } y_i = y_j \\ 1 & \text{if } y_i \neq y_j, \end{cases} \tag{5.31}$$

a social polarization index is defined to be proportional to

$$\mathrm{DP}(\alpha, \mathbf{F}) = \sum_{i=1}^{n}\sum_{j\neq i}^{n}\pi_i^{1+\alpha}\pi_j \tag{5.32}$$

with $\alpha \in [0, \infty[$. Alienation is thus set to zero between individuals belonging to the same group and to one between individuals belonging to different social groups. Each value of α leads to a different social polarization index. Restricting the range of these values can be done using two axioms reminiscent of those in Esteban and Ray (1994).

Axiom MRQ 1

Let there be three population groups of sizes p, q, and r where $p > q \geq r$. If the two smaller groups are merged to form one new group with size $\widetilde{q} = q + r$, then polarization increases.

Montalvo and Reynal-Querol (2002) (see also Montalvo and Reynal-Querol, 2005a,b, 2008) show that the social polarization index $\mathrm{DP}(\alpha, \mathbf{F})$ satisfies Axiom MRQ 1 if and only if $\alpha \geq 1$. The importance of the identification effect must indeed be sufficiently strong. The impact of the disappearance of one group reduces the number of pairs of alienated groups; for $\alpha \geq 1$, this is offset by the identification effect of the formation of a larger group.

Axiom MRQ 2

Let the population be split into three groups of sizes p, q, and p. If there is an equal movement of the population from group q to the two other groups, then polarization increases.

The only social polarization index $\mathrm{DP}(\alpha, \mathbf{F})$ that satisfies Axiom MRQ 2 is the one with $\alpha = 1$. $\mathrm{DP}(1, \mathbf{F})$ is the Reynal-Querol (2002) polarization index:

$$\mathrm{RQ} = 1 - \sum_{i=1}^{n} \frac{(0.5 - \pi_i)^2 \pi_i}{0.25} = 4 \sum_{i=1}^{n} \sum_{j \neq i}^{n} \pi_i^2 \pi_j. \tag{5.33}$$

As shown in Equation (5.33), the RQ index was originally motivated and interpreted as one minus the average of the squared distances of group sizes from 0.5. In that formulation, the simple distribution of two equal-size groups is implicitly considered as the one with the most social polarization. Taking the square of the distances of group sizes from 0.5 captured the gap between group sizes and the symmetric two-group distribution. An increase in the number of groups or spreads of group sizes from 0.5 leads to lower social polarization.

Apart from fulfilling Axioms MRQ 1 and MRQ 2, the RQ index has two additional interesting properties. With four groups, minimum social polarization is obtained when groups are all of the same sizes. Second, with n groups of equal relative size $1/n$, social polarization decreases monotonically with n (see Figure 5.19).

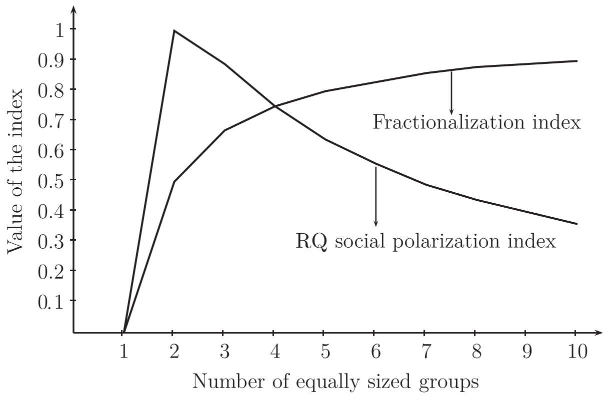


Figure 5.19 Fractionalization (FRAC) and social polarization (RQ) indices as a function of a number of equally sized groups. *Source: Montalvo and Reynal-Querol (2008).*

Chakravarty and Maharaj (2011b) also characterized the Reynal-Querol (2002) index of social polarization using alternative sets of independent axioms. Chakravarty and Maharaj (2012) further showed how a graph of cumulative population shares against the cumulative number of groups (ranked from the largest to the smallest ones) can rank some classes of social polarization indices of an RQ type and propose in the process a "generalized RQ index of order θ" given by

$$\mathrm{RQ}(\theta) = 4\sum_{i=1}^{n} \pi_i^2 (1-\pi_i) + \theta \sum_{1\le i_1<i_2<i_3\le n} \pi_{i_1}\pi_{i_2}\pi_{i_3}, \tag{5.34}$$

where $\theta \in [0,3]$ and $n \geq 3$. RQ(0) is the usual RQ index.

Much of the literature prior to Montalvo and Reynal-Querol (2002) has attempted to explain conflict through a social (often an ethnolinguistic) fragmentation index of the form:

$$\mathrm{FRAC} = 1 - \sum_{i=1}^{n} \pi_i^2. \tag{5.35}$$

The FRAC index, which is the complement to unity of the Hirschman–Herfindahl index (see Herfindahl, 1950; Hirschman, 1980), can be interpreted as the probability that two persons randomly drawn from a given society will not belong to the same social group. The probability that two randomly drawn individuals belong to the same group i is indeed π_i^2; one minus the sum of those probabilities across all groups gives FRAC.

The fractionalization index can also be derived from the Gini index, which is proportional to $\mathrm{ER}(0, \mathbf{F})$ in Equation (5.29). Replacing the Euclidean distance $|y_i - y_j|$ by the discrete distance in Equation (5.31), we obtain

$$\mathrm{DP}(1, \mathbf{F}) = \sum_{i=1}^{n} \sum_{j\neq i}^{n} \pi_i \pi_j = 1 - \sum_{i=1}^{n} \pi_i^2 = \mathrm{FRAC}. \tag{5.36}$$

Which of FRAC or of RQ should be a better predictor of conflicts has been a matter of debate. The distinction between the two is analogous to that between inequality and polarization. Horowitz (1985) argued that there should be less violence in highly homogeneous and highly heterogeneous societies and more conflict in societies in which a large ethnic minority coexists with an ethnic majority. An increase in social heterogeneity should initially increase potential conflict, but after some point, greater diversity should eventually imply lower potential for conflicts—see also Colomer (2001) and Collier and Hoeffler (2004). If this is correct, then an index of social polarization should capture the likelihood of conflict, or the intensity of potential conflict, better than an index of fractionalization.

The empirical literature provides some support to this view. Collier and Hoeffler (2004) found that religious fractionalization has no effect on the risk of conflict.

Montalvo and Reynal-Querol (2003, 2008) found little empirical evidence of an effect of fractionalization on growth, but did report an important effect of religious and ethnic polarization on civil wars, investment, and the share of government consumption out of GDP, and thus on growth. Evidence showing that fractionalization is a better predictor of conflict can nevertheless be found in Alesina et al. (2003).

One way to understand better the distinction between indices of social polarization and indices of fractionalization is to consider two populations A and B, each composed of three different groups (this example is discussed in Montalvo and Reynal-Querol, 2002). The sizes of the groups in population A are given by (0.49, 0.49, and 0.01) and the sizes in population B are given by (0.33, 0.33, and 0.34). Which of these two populations is more likely to witness conflict? Fractionalization is higher in B; social polarization is larger in A. The likelihood of two individuals belonging to two different groups is indeed larger in B, but A is closer to a distribution with two equal-size groups, and thus closer to maximal social polarization.

The index of fractionalization increases whenever two-group sizes are equalized; the RQ of social polarization index increases only when these two groups are relatively large ones. The value of α in $\mathrm{DP}(\alpha, \mathbf{F})$ cannot exceed 1, however, because with $\alpha > 1$ $\mathrm{DP}(\alpha, \mathbf{F})$ does not monotonically increase in moving from (0, 1, and 0) toward (0.5, 0, and 0.5)—and $\mathrm{DP}(\alpha, \mathbf{F})$ thus fails Axiom MRQ 2 when $\alpha > 1$.

The issue echoes the discussion of Axiom MRQ 2 on given earlier. If we shift population mass from one group equally to two other groups of equal size, then the axiom says that social polarization should increase. As mentioned earlier, the only admissible value of α for a $\mathrm{DP}(\alpha, \mathbf{F})$ measure to satisfy that axiom is 1. With $\alpha = 0$ in particular, fractionalization is not maximized with a distribution of two equal-sized groups; in fact, FRAC increases monotonically with n when n is the number of groups with equal relative sizes given by $1/n$ (Figure 5.19).

5.7. SOCIOEONOMIC POLARIZATION

5.7.1 Between- and Within-Group Income Inequality

The inequality literature has long used social characteristics to decompose income inequality. Based on this, Zhang and Kanbur (2001) suggested using that within-group inequality to capture internal heterogeneity and between-group inequality to measure external heterogeneity. They proposed the ratio of between-group inequality to within-group inequality as the following socioeconomic polarization index[13]:

[13] Chakravarty and Maharaj (2011a) considered generalized forms and orderings of polarization indices based on such between- and within-group inequality terms—see also Chakravarty et al. (2010).

$$\text{ZK} = \frac{\text{between-group inequality}}{\text{within-group inequality}}. \tag{5.37}$$

Obvious candidates of inequality indices for this kind of exercise are the generalized entropy indices that are well known to be exhaustively decomposable into between- and within-group components.[14] As within-group income differences decline, that is, as the groups become internally more equal (though not necessarily more identified, see the following discussion), differences across groups are, relatively speaking, magnified, and polarization rises. For given within-group differences, however, polarization increases as the group means drift apart and between-group inequality increases.

The use of within-group inequality to capture internal homogeneity is not compatible with the identification framework in the income polarization literature. To see why, consider Figure 5.18, which was discussed earlier. Think of two groups each being formed by two subgroups on the same side of the median line. As shown in the figure, a split of the two right-hand side groups increases internal homogeneity and therefore decreases within-group inequality. According to Equation (5.37), that should increase polarization, though this is not what the identification/alienation framework would necessarily conclude.

Another feature of socioeconomic polarization is that it may conflict with income polarization and income bipolarization, in part because socioeconomic polarization does not use income to form social groups. The presence of identical groups can lead to indices such as ZK to show no socioeconomic polarization, although the income distribution may exhibit any level of income polarization and bipolarization. In Figure 5.23, for instance, there may be no socioeconomic polarization, but substantial income polarization.

5.7.2 Identification/Alienation Hybrids

Several hybrids of the identification/alienation indices can be accommodated, using the idea that identification can be mediated not only by group membership, but also by income similarities as well, whereas the antagonism function remains defined on income differences. Duclos et al. (2004) proposed, for instance, a "social polarization index with income-mediated identification":

$$\text{DER}_{\text{H1}} = \sum_{i=1}^{n} (1 - n_i) \int_x f_i(y)^{\alpha} \mathrm{d}F_i(y) \tag{5.38}$$

where α is the usual parameter of polarization sensitivity.

[14] Deutsch and Silber (2010) used between- and within-group components of the Gini index to measure socioeconomic polarization.

Another hybrid considers alienation as well as identification in terms of income mediation; two individuals must belong to different groups and have different incomes to experience alienation. The polarization measure that is obtained is:

$$\mathrm{DER}_{\mathrm{H2}} = \sum_{j=1}^{n}\sum_{k\neq j}^{n}\int_x\int_y f_j(x)^{\alpha}|x-y|\mathrm{d}F_j(x)\mathrm{d}F_k(y). \tag{5.39}$$

Alienation uses income distances; identification is based on income and social characteristics alike.

Permanyer (2010) presented two axiomatically characterized socioeconomic polarization indices incorporating a variable alienation component. The focus is first on between-group alienation only and subsequently on within-group alienation as well. The starting point is the identification–alienation framework.

Assume that there are n exogenous social groups (based on religious, ethnic, or political characteristics, for instance) that can possibly inform an individual's sense of identity. Each individual also feels a degree of "radicalism" $y > 0$. Radicalism is defined by the degree with which an individual defends the identity/interests/objectives of his or her group. Radicalism serves two purposes. First, y measures the strength with which individuals compare themselves with others in different social groups. Second, y can fuel an individual's sense of identity/difference toward others in the same social group.

Thus, there are two potential sources of identification and two potential bases for alienation. Identification can depend either solely on the size of one's social group or on both the size of that group and on the degree of radicalism felt by individuals within that group. Alienation is felt toward those members of the other social groups, with an intensity given by the sum of the radicalism felt by individuals of different groups; different degrees of radicalism among individuals of the same group can also fuel alienation across members of that same group.

This setting is a departure from the income polarization and social polarization frameworks. Unlike income polarization, identity can be a function both of income and of membership into a social group, and there may or may not be alienation between members of a same group. Unlike social polarization, identity can be fueled by the degree to which members associate with the interests of their group, and not only by pure membership in the group; the intensity of alienation can also be permeated by that same degree of radicalism.

The alienation–identification framework is then adopted in Permanyer (2010), first by assuming that there is no within-group alienation. Identity depends only on group size n_i. Individuals feel perfectly identified with all the members of their own group—the groups are cohesive and membership into them is sufficient for defining individuals' sense of identity. Alienation across members of two different groups with radicalism x and y is defined as being a monotonically increasing function of the sum of radicalism, $x + y$. The higher the force with which individuals defend the interests of their group, the higher the animosity felt toward individuals of a different group; the sum of radicalism thus appears as a measure of tension between individuals of different groups. This is a departure from the usual alienation framework, in which distances (and not sums) capture alienation.

Using the expressions for alienation and identification, we then have

$$\mathrm{Per}(\mathbf{f}) = \sum_{i=1}^{n} \sum_{j \neq i}^{n} \int\int T(n_i, x+y) f_i(x) f_j(y) \mathrm{d}y \mathrm{d}x. \tag{5.40}$$

The rest of the analysis follows broadly the axiomatic framework of Duclos et al. (2004). Three axioms on movements of densities within and across groups and an additional population invariance axiom are imposed. Figure 5.20 illustrates the first axiom.

Consider a given group divided into two subgroups, the bigger one with a lesser degree of radicalism and the smaller subgroup with a greater degree of it. If these subgroups respectively increase and decrease their degrees of radicalism by the same amount, then the average degree of radicalism within the group increases. The first axiom (a within-group axiom) says that this should raise overall social economic polarization. It also implies that T is concave in $x+y$.

The second axiom is based on the net effect of a smaller group becoming less radical and of a bigger group becoming more radical. This is shown in Figure 5.21. This second axiom (a between-group axiom) says that the effect of the increase in radicalism in the bigger group should dominate the impact of the fall of radicalism in the second group; polarization should not decrease. It also implies that T is convex in $x+y$.

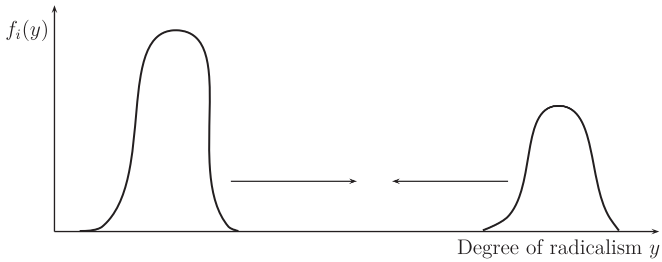


Figure 5.20 A slide of basic densities within a group increases socioeconomic polarization.

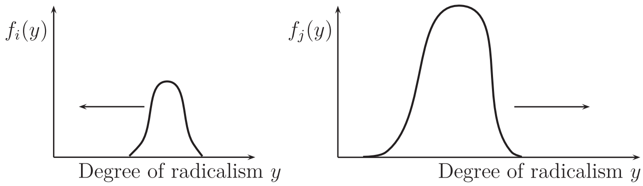


Figure 5.21 A smaller group becoming less radical and a bigger group becoming more radical does not decrease socioeconomic polarization.

The third axiom says that a movement of population from a large group to two equally smaller groups, with the same normalized density, should not decrease polarization. This implies that the identification effect should not be too large (and therefore imposes and upper bound on α in Equation (5.41)). An equalization of population sizes across these three groups with the same distribution of radicalism will thus generate more socioeconomic polarization than if one group dominates size. Added to the usual population invariance property, this necessarily leads to the index:

$$\mathrm{Per}(\alpha, \mathbf{F}) = \sum_{i=1}^{n} \sum_{j \neq i}^{n} \pi_i^{1+\alpha} \pi_j \left(\mu_i + \mu_j \right), \tag{5.41}$$

where $\alpha \in (0, 1]$. This measure of socioeconomic polarization generalizes the social polarization $\mathrm{DP}(\alpha, \mathbf{F})$ (and thus also the RQ index). Instead of setting alienation between groups to a constant, it is made sensitive to the size of radicalism and to different possible levels of animosity between groups. If the means μ_i are identical for all groups, then $\mathrm{Per}(\alpha, \mathbf{F})$ simplifies to a proportion of $\mathrm{DP}(\alpha, \mathbf{F})$. In the particular case of $\alpha = 0$, $\mathrm{Per}(\alpha, \mathbf{F})$ is generalization of the fractionalization index FRAC, and it simplifies to the classical fractionalization index when $\mu_i = \mu_j, \forall\, i \neq j$. The value of α has the same interpretation as the polarization sensitivity parameter in Esteban and Ray (1994); the larger it is, the greater the departure from (socioeconomic) inequality.

The lower bound of α in Equation (5.41) can be set above zero if we require polarization to fall as a number n of identical groups (with relative size $1/n$) increases, thus differentiating socioeconomic polarization from fractionalization and inequality. Permanyer (2010) then showed that the lower bound of α becomes $(2 - \log_2 3)/(\log_2 3 - 1) = 0.71$.

The preceding ignores within-group alienation in socioeconomic polarization. Alienation may exist between members of the same group if levels of radicalism differ; within a group, radical members may alienate the more moderate members and vice versa. Permanyer (2010) adapted this by letting between-group alienation be measured by a monotonically increasing function in $x + y$, but by supposing within-group alienation to be measured by a monotonically increasing function in $|x - y|$, as is usual in the identification/alienation framework. Under this setting, total polarization becomes

$$\begin{aligned} \mathrm{Per}(\mathbf{f}) = {} & \sum_{i=1}^{n} \int\!\!\int T(f_i(x), |x - y|) f_i(x) f_i(y)\, \mathrm{d}y\, \mathrm{d}x \\ & + \sum_{i=1}^{n} \sum_{j \neq i}^{n} \int\!\!\int T(f_i(x), x + y) f_i(x) f_j(y)\, \mathrm{d}y\, \mathrm{d}x. \end{aligned} \tag{5.42}$$

The first component on the right-hand side of Equation (5.42) represents the contribution of within-group polarization and the second component is the contribution of between-group polarization. Within- and between-group components are summed to obtain total polarization.

Using population invariance axioms similar to Axioms DER 2 and DER 3 and a "socioeconomic polarization axiom" that says that a population transfer across two initially identical groups should lower polarization, the index becomes

$$\mathrm{Per}(\alpha, \mathbf{f}) \equiv \sum_{i=1}^{n} \int\int f_i^{1+\alpha}(x) f_i(y)|x-y|\mathrm{d}y\mathrm{d}x + \sum_{i=1}^{n}\sum_{j\neq i}^{n} \int\int f_i^{1+\alpha}(x) f_j(y)(x+y)\mathrm{d}y\mathrm{d}x \tag{5.43}$$

with $\alpha \in [1/(3n-2), 1]$, or with $\alpha \in [0.5, 1]$ if we require polarization to fall as a number of identical groups increases.

Social polarization and income polarization do not deal with situations in which a society may be partitioned with respect to several variables, some of a social and others of an economic type. Consider, for instance, the case in which individuals may be male or female and may enjoy different health statuses. This is illustrated in Figures 5.22 and 5.23. In the usual social and income polarization settings, the situation in which all males (represented by the dark areas) have very poor heath and all females (represented by the white rectangles) have very good heath (Figure 5.22) leads to the same degree of polarization (and bipolarization) as the one (Figure 5.23) in which half of males and half of females have very poor heath and the other halves very good health. This is because income polarization does not take into account the effect of segregation by social

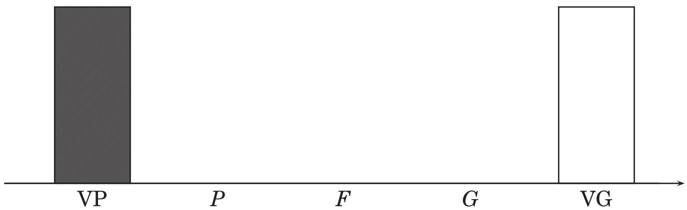


Figure 5.22 A hypothetical distribution of health statuses across two different social groups (males and females)—the initials stand for: very poor (VP), poor (P), fair (F), good (G), and very good (VG). *Source: Permanyer and D'Ambrosio (2013).*

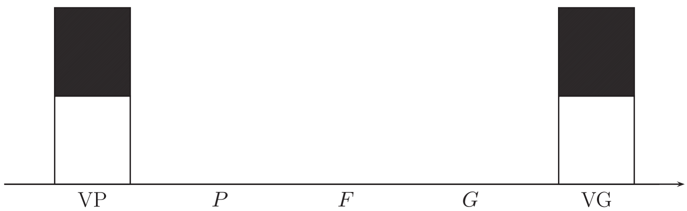


Figure 5.23 A hypothetical distribution of health statuses across two different social groups (males and females)—the initials stand for: very poor (VP), poor (P), fair (F), good (G), and Very Good (VG). *Source: Permanyer and D'Ambrosio (2013).*

characteristics and because social polarization does not consider distances in welfare across groups. It seems plausible, however, that the situation in Figure 5.22 may exhibit more tension and more polarization than the second, in which males and females are mixed in health status. Polarization should therefore presumably be sensitive to the joint distribution of social groups and of welfare statuses.

Permanyer and D'Ambrosio (2013) pursued this issue by first partitioning the population of individuals into exogenously given groups. Members of a given group identify with their peers within their group but feel alienated from all others. Identification for members of a particular group i depends on the size of the group to which they belong (n_i). Identification is a function only of group sizes. Alienation is assumed to be the same for every member of a particular group; between-group rather than between-individual alienation matters.

Alienation is captured by an overlap measure. This is different from the usual distance-based definition of alienation. Let $\theta_{i,j}$ be an overlap coefficient between groups i and j in a context of categorical welfare data, with C categories, and where $\pi_i(y_c)$ is the proportion of individuals in group i that have a level of welfare equal to y_c:

$$\theta_{i,j} = \sum_{c=1}^{C} \min\left\{\pi_i(y_c), \pi_j(y_c)\right\}. \tag{5.44}$$

Alternatively, with a continuous welfare variable y, we have:

$$\theta_{i,j} = \int \min\left\{\bar{f}_i(y), \bar{f}_j(y)\right\} \mathrm{d}y, \tag{5.45}$$

where $\bar{f}_i(y)$ is the normalized density of group i over welfare y. Hence, in capturing alienation, it is not distances that matter but rather the importance of the clustering of groups in certain areas of a welfare domain. This is different from social and income polarization. By Equations (5.44) and (5.45), $0 \leq \theta_{i,j} \leq 1$. When groups are disjoint, $\theta_{i,j} = 0$, and when groups overlap perfectly, $\theta_{i,j} = 1$. Alienation is defined by $1 - \theta_{i,j}$, which equals 1 in the case of completely disjoint groups and 0 in the case of perfectly overlapping groups.

Following the identification–alienation framework of Esteban and Ray (1994), polarization can be defined as the sum of all effective antagonisms, that is, as

$$\mathrm{PD_S} = \sum_{i=1}^{n} \sum_{j=1}^{n} n_i n_j T\left(n_i, 1 - \theta_{ij}\right). \tag{5.46}$$

Using axioms similar to Duclos et al. (2004) but adapted to a socioeconomic setting with multiple social groups, Permanyer and D'Ambrosio (2013) found that a socioeconomic polarization index of the form Equation (5.46) should be proportional to

$$\mathrm{PD_S} \equiv \sum_{i=1}^{n} \sum_{j=1}^{n} \pi_i^{1+\alpha} \pi_j \left(1 - \theta_{ij}\right), \tag{5.47}$$

where the degree of polarization sensitivity α lies within $[\alpha^*, 1]$, with $\alpha^* = (2 - \log_2 3)/(\log_2 3 - 1) = 0.71$.

A socioeconomic extension of Esteban et al.'s (2007) income polarization index is provided in Gradín (2000). The extension uses both income distribution and social differences to form groups, in contrast to the $\mathrm{EGR}(\alpha, \sigma, \mathbf{F})$ for which groups are determined on the basis of income differentials. The objective is essentially to assess by how much social polarization correlates with income polarization; the greater the correlation, the greater the polarization association between income and social characteristics.

The first socioeconomic polarization index (called "group polarization") considers a population of n subgroups defined on the basis of social characteristics, such as education level, skin color, gender, race, region, and so on, rather than income, as for the ER and EGR indices. The socioeconomic grouping induces a partitioning of the income distribution given by $\mathbf{F}^{\mathrm{g}}$. This leads to a socioeconomic polarization index given by

$$\mathrm{EGR}(\alpha, \sigma, \mathbf{F}^{\mathrm{g}}) = \mathrm{ER}(\alpha, \mathbf{F}^{\mathrm{g}}) - \sigma\epsilon(\mathbf{F}^{\mathrm{g}}). \tag{5.48}$$

This social categorization of the population leads to lower between-group income dispersion and higher within-group heterogeneity than the "optimal" EGR income partitioning $\mathbf{F}^*$. $\mathrm{EGR}(\alpha, \sigma, \mathbf{F}^{\mathrm{g}})$ is thus expected to be lower than $\mathrm{EGR}(\alpha, \sigma, \mathbf{F}^*)$. The lower the difference between these two expressions, the greater the ability of social groupings to explain income polarization.

Gradín (2000) also proposed an alternative measure of the association between income polarization and socioeconomic polarization (called "explained polarization"). To see how, consider $n+1$ income cutoffs $z = z_0, z_1, \ldots, z_n$ and the set of groups $\phi_j = \{i | m_i \in [z_{j-1}, z_j]\}$ (for $j = 1, \ldots, n$) made of those social groups i whose average income lies in $[z_{j-1}, z_j]$. $\mathbf{F}^{\mathrm{e}}$ then has elements given by $F_j^{\mathrm{e}}(y) = \sum_{i \in \phi_j} F_i(y)$. $\mathbf{F}$ is a partition of the population according to social characteristics $i = 1, \ldots, n$; $\mathbf{F}^{\mathrm{e}}$ is a repartition of that partition that reclassifies individuals according to the mean income of the group to which they belong along the thresholds $z = z_0, z_1, \ldots, z_n$; and the movement from $\mathbf{F}$ to $\mathbf{F}^{\mathrm{e}}$ is an economic reordering of social groups. $\mathrm{EGR}(\alpha, \sigma, \mathbf{F}^{\mathrm{e}})$ is then the level of polarization with socioeconomic grouping $\mathbf{F}^{\mathrm{e}}$.

Given this, we may wish to compute how much of income polarization can be explained by $\mathbf{F}^{\mathrm{e}}$. Gradín (2000) did this by defining a ratio of explained polarization given by:

$$\mathrm{EP}(\alpha, \beta, \mathbf{F}, \mathbf{F}^{\mathrm{e}}) = \frac{\mathrm{EGR}(\alpha, \beta, \mathbf{F}^{\mathrm{e}}) - \mathrm{EM}(\mathbf{F}^{\mathrm{e}})}{\mathrm{EGR}(\alpha, \beta, \mathbf{F}) - \mathrm{EM}(\mathbf{F})} \tag{5.49}$$

where EM is EGR when there is no between-group heterogeneity.

5.8. MULTIDIMENSIONAL POLARIZATION

Gigliarano and Mosler (2009) constructed multivariate socioeconomic polarization indices on the presumption that polarization should capture internal homogeneity, external heterogeneity, and similarity of group sizes. Compared to Zhang and Kanbur (2001), the measurement of group homogeneity and heterogeneity is refined by using multivariate distances. Multivariate decomposability into intragroup and intergroup components of inequality is then used to construct multidimensional polarization indices.

To see how, consider a distribution of N individuals along C types of endowments, represented by a matrix $\mathbf{X}$

$$\mathbf{X}=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1C} \\ x_{21} & x_{22} & \cdots & x_{2C} \\ \vdots & & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NC} \end{bmatrix}_{N\times C}$$

where x_{iq} denotes the endowment q of individual i. Using inequality decomposition, the multidimensional polarization index is given by:

$$\mathrm{GM}(\mathbf{X})=\zeta[\mathrm{GB}(\mathbf{X}),\mathrm{GW}(\mathbf{X}),S(\mathbf{X})], \tag{5.50}$$

where GB and GW are multivariate indices that measure inequality between and within groups, respectively, and where ζ is a function that is increasing in GB and S and is decreasing in GW. $S(\mathbf{X})$ is a measure of deviation from equally sized groups; it is maximal if all groups are of the same size. Particular forms of this index are given by

$$\mathrm{GM}_1(\mathbf{X})=\phi\left[\frac{\mathrm{GB}(\mathbf{X})}{\mathrm{GW}(\mathbf{X})+c}\right]S(\mathbf{X}), \tag{5.51}$$

$$\mathrm{GM}_2(\mathbf{X})=\psi[\mathrm{GB}(\mathbf{X})-\mathrm{GW}(\mathbf{X})]S(\mathbf{X}), \tag{5.52}$$

$$\mathrm{GM}_3(\mathbf{X})=\tau\left[\frac{\mathrm{GB}(\mathbf{X})}{\mathrm{GB}(\mathbf{X})+\mathrm{GW}(\mathbf{X})+c}\right]S(\mathbf{X}), \tag{5.53}$$

with constant c being positive and possibly dependent on the choice of indices GB and GW The functions ϕ, ψ, and τ are assumed to be continuous and strictly increasing, with $\phi(0)=\tau(0)=0$. These forms were subsequently developed by Gigliarano and Mosler (2009) using additively decomposable multivariate inequality measures. Similar procedures are proposed for multiplicatively decomposable inequality measures. Equation (5.51) is similar to Zhang and Kanbur (2001) except for the explicit role of group sizes $S(\mathbf{X})$ and for the fact that $\mathbf{X}$ is a matrix. Equation (5.52) is similar to the Foster and Wolfson (2010/1992) bipolarization measure with below and above median groups—although the definition of a "median frontier" in a multidimensional setting is not immediate.

Two alternative multivariate polarization indices with special relevance to the case of a population split into two groups are proposed in Anderson (2010). When these two groups are made of poor and nonpoor individuals, these indices can be used as multivariate relative poverty measures based on distances between them. Consider then a population composed of two groups: the poor and the nonpoor. They are assumed to be distributed according to two continuous multivariate unimodal distributions: $f_{\mathrm{p}}(\mathbf{x})$ for the poor and $f_{\mathrm{r}}(\mathbf{x})$ for the nonpoor, where $\mathbf{x}$ is a $1 \times C$ vector of characteristics. An "overlap" measure is defined as:

$$\mathrm{OV} = \int_{\mathbf{x}} \min \left\{ f_{\mathrm{p}}(\mathbf{x}) f_{\mathrm{r}}(\mathbf{x}) \right\} \mathrm{d}\mathbf{x}. \tag{5.54}$$

This is a multidimensional extension of the overlap measure of Equation (5.45). Assessing the degree of polarization between two groups when individuals have many characteristics amounts in this formulation to capturing the degree of commonality between the two distributions.

Anderson (2010) proposed an alternative index that can be used when attributes are mutually exclusive and do not overlap. Let $\mathbf{x}_{\mathrm{mp}}$ and $\mathbf{x}_{\mathrm{mr}}$ be the value of the characteristic vector at the modal point of the poor and the nonpoor distributions, respectively. The index is the area of the trapezoid formed by the modal peaks of the densities and the mean normalized Euclidean distance between these two points. Let μ_q be the mean of the q th characteristic in the pooled population. When the poor and nonpoor distributions are separately identified in C dimensions, multidimensional polarization is written as:

$$\mathrm{BIPOL_I} = 0.5\left[f_{\mathrm{p}}\left(\mathbf{x}_{\mathrm{mp}}\right) + f_{\mathrm{r}}\left(\mathbf{x}_{\mathrm{mr}}\right)\right] \frac{1}{\sqrt{C}} \sqrt{\sum_{q=1}^{C} \frac{\left(x_{\mathrm{mpq}} - x_{\mathrm{mrq}}\right)^2}{\mu_q}}. \tag{5.55}$$

This can be interpreted as a multivariate relative poverty index without formal poverty frontiers, the index being based on distances between two identified groups, poor and nonpoor. Note, however, that two distributions of the poor and nonpoor may not be separately identified, in which case it may be difficult to apply an index like Equation (5.55). Unique modal distances may also not exist between the poor and the nonpoor. Even if they exist, it is not clear that they would necessarily be appropriate measures of deprivation over the entire distributions: Many different multivariate distributions can exhibit similar distances between such modes but may yet represent quite different allocations of welfare across the poor and the nonpoor.

A multidimensional extension of DER's identification/alienation framework can also be designed. A natural procedure for that is found in Anderson (2011). The extension is multidimensional in the sense that both identification and alienation depend on the joint distribution of multiple socioeconomic attributes. The multivariate distribution can also be defined over a collection of both discrete and continuous variables.

To see how, let $\mathbf{u}_i$ and $\mathbf{v}_j$ be stacked vectors of continuous and discrete variables, with dimension k (for the continuous variables) and h (for the discrete variables) and for individuals i and j, respectively. A multivariate polarization index taking into account both continuous and discrete variables is then given by

$$\mathrm{AN}(\alpha)=\int_u\int_v [f(u)^{\alpha}||\mathbf{u}-\mathbf{v}||\mathrm{d}F(\mathbf{u})\mathrm{d}F(\mathbf{v}), \tag{5.56}$$

where $||.||$ represents a C-dimensional Euclidean norm defined by

$$||\mathbf{u}_i-\mathbf{v}_j||=\sqrt{\frac{C\sum_{q=1}\left(u_{iq}-v_{jq}\right)^2}{C}}. \tag{5.57}$$

This is the normalized Euclidean distance between individuals i and j, where u_{iq} is the q th variable of individual i and $C=k+h$.

5.9. POLARIZATION IN PRACTICE

A few empirical studies on polarization have been conducted both to illustrate the use of the various indices that have been proposed and to motivate the study of polarization (by looking at the links between polarization, inequality, and conflict, for instance). We end this chapter with brief remarks on such empirical evidence: first, on the links between polarization and inequality; second, on polarization comparisons; and finally, on the linkages between polarization and conflict.

5.9.1 Polarization and Inequality

Wolfson (1994) empirically investigated the divergence between inequality and bipolarization using time-series data from the Canadian Surveys of Consumer Finance from 1967 to 1991. Although inequality and polarization often move in the same direction, from 1973 to 1981 they diverged.

Chakravarty and Majumder (2001) used Indian household expenditure survey data for 1987–1988 and 1993–1994 to compare inequality and polarization across time, states, and rural and urban areas. In some states in rural India, overall inequality decreased. The inequality ranking of states is unchanged between the two periods, but polarization rankings do vary. In some states, polarization and inequality moved in opposite directions regardless of polarization parameter values, reinforcing the view that polarization and inequality can behave differently.

Using data come from 28 Chinese provinces between 1983 and 1995, Zhang and Kanbur (2001) found that, contrary to other earlier empirical evidence, polarization measures do not generate significantly different results from standard measures of inequality. Polarization and inequality measures agree broadly over the period considered. This motivates them to consider a measure of socioeconomic polarization ZK, which they find empirically useful to examine the case of the United States, where socioeconomic polarization appears to exist between blacks and whites, and the case of China, where differences between rural and urban as well as between coastal and inland groups are important.

Some of the most extensive empirical evidence on the association between inequality and polarization can be found in Duclos et al. (2004). DER analyze polarization using several waves of LIS data from 21 countries. Their empirical analysis shows that polarization is empirically different from inequality. DER($\alpha = 0$) and the Gini index are by definition identical. When $\alpha = 0.25$, the Gini coefficient and the polarization index yield similar rankings. The correlation between inequality and polarization rankings falls when the polarization parameter α increases. Much of the reordering of countries as α increases to 1 occurs at low values of α. Rank differences between the Gini coefficient and the polarization index are important when $\alpha = 1$, with a Pearson rank correlation coefficient often lower than 0.7.

Figure 5.24 shows the Duclos et al. (2004) estimates for $\alpha = 1$ and for the 21 countries of LIS Wave 3. For these countries, a linear regression of inequality on polarization yields an R^2 of 0.8—understandably large but not perfect. The Czech Republic has the lowest Gini index of all countries, but ranks 11 out of 21 in terms of polarization. Conversely, Canada, Australia, and the United States exhibit large inequality, but relatively low polarization. The bulk of cross-country variation in polarization comes both from a significant variation in average identification and from variability in average alienation (and not from the covariance between identification and alienation—recall the discussion at the end of section 4.2); this explains why inequality and polarization are correlated but still empirically quite different. The United Kingdom and the United States also provide an interesting comparison. Both have a similar level of average

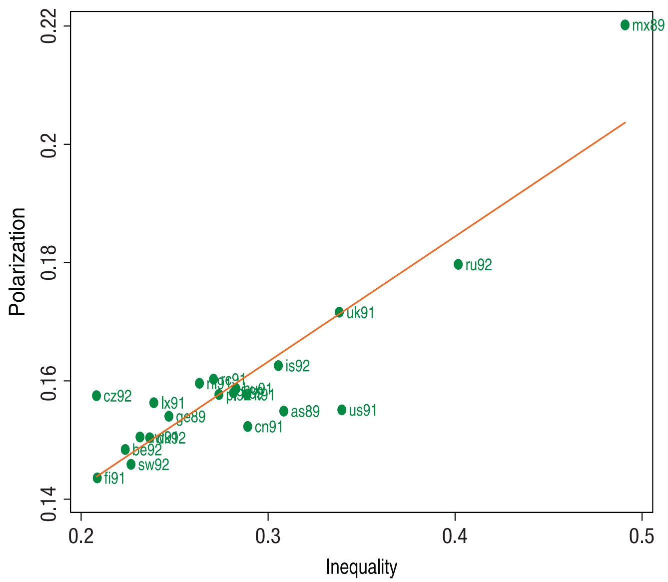


Figure 5.24 Polarization (measured by the DER ($\alpha = 1$) index) and inequality for 21 countries, LIS Wave 3. *Data source:* Duclos et al. (2004).

alienation, but the United Kingdom's density has sharper peaks, with the United States' density being remarkably flat with thick tails. The United Kingdom then ends up with a higher level of polarization, especially with higher values of α.

5.9.2 Empirical Polarization Comparisons

Using their range-free dominance techniques, Foster and Wolfson (2010/1992) compared bipolarization in the United States and Canada between 1979/1981 and 1988. Polarization and inequality are increasing in the United States but stable in Canada. The United States is robustly more polarized and unequal than Canada.

Duclos and Échevin (2005) applied their first-order bipolarization ordering techniques to rank 29 countries. They established bipolarization comparisons for 32% of the pairs of countries. This percentage increases to 57% when orderings are established solely on sample estimates and statistical uncertainty is ignored. Mexico has the highest level of bipolarization of all 29 countries, the ranking being statistically significant with respect to 26 countries, followed by Russia and the United States.

The extended polarization measure developed by Esteban et al. (2007) is applied to the income distribution of five OECD countries, namely the United States, the United Kingdom, Canada, Germany, and Sweden for the period 1974–2000. The main temporally observed pattern is a decrease in the Esteban and Ray (1994) polarization index combined with a moderate decline in within-group dispersion. Because the Esteban et al. (2007) index is a difference between these two factors, the net pattern is a decrease in the EGR index, especially in Canada and Germany. In 1974, at the beginning of the period, only Sweden presented relatively low polarization; Germany and the United Kingdom had similar levels of polarization, being nearly as polarized as Canada and the United States. By 2000, the United States and the United Kingdom were highly polarized relative to Sweden, Germany, and Canada.

Using the intermediate polarization measure, Chakravarty and D'Ambrosio (2010) compared polarization across four southern European countries, Greece, Italy, Spain, and Portugal. With regard to absolute polarization, Italy is the most polarized country on average, followed by Spain. But with respect to relative and intermediate polarization, Italy is the least polarized country on average, whereas Portugal is the country showing the highest level of polarization.

Anderson (2011) combined the indices of Esteban and Ray (1994) and Duclos et al. (2004) to consider polarization across discrete and continuous variables. Unidimensional polarization is found to have increased between 1990 and 2001 in the six Chinese provincial distributions of urban households considered, over the dimensions of income, living space, and education separately. When the three variables are combined in a multidimensional space, polarization is stable over 1990–2001 for smaller values of polarization aversion and decreasing for higher values.

5.9.3 Polarization and Conflict

Reynal-Querol (2002) estimated the impact of religious polarization and animist diversity on the incidence of ethnic civil war. Her main finding is that religious divisions seem more important than language divisions and natural resources in explaining social ethnic conflicts. She found in particular a positive and robust impact of animist diversity and religious polarization (but not of language fractionalization) on the incidence of ethnic civil war.

The effect of polarization and fractionalization on the incidence of civil and ethnic wars and, indirectly, on economic growth was further studied in Montalvo and Reynal-Querol (2005a). No relationship between fractionalization and conflict emerges for more than two groups. Ethnic polarization has a positive and significant impact on the prevalence of civil wars, as does religious polarization. Fractionalization and polarization are even negatively correlated at higher levels of fractionalization.

Montalvo and Reynal-Querol (2003) further support the evidence (reported in Montalvo and Reynal-Querol, 2002) that religious polarization is more appropriate for measuring the effect of potential conflict on economic development than the traditional fractionalization index. Religious polarization has a negative and significant effect on development through its effect on investment, government expenditure, and the frequency of civil wars. No effect of religious and ethnic fractionalization is found on the incidence of ethnic and civil wars on growth or on investment.

In an extension to the preceding, Montalvo and Reynal-Querol (2005b) reported that ethnolinguistic fractionalization has no statistically significant effect on the incidence of civil war, but that ethnic polarization does have a positive and statistically significant impact. Montalvo and Reynal-Querol (2008) further found that polarization has a positive and significant effect on genocide, whereas fractionalization does not.

Unlike social polarization indices such as that of Reynal-Querol (2002), the socioeconomic polarization index of Permanyer (2010) takes into account the extent of alienation between individuals. Using data from the World Value Surveys on 65 countries, the indices of Reynal-Querol (2002) and Permanyer (2010) are found to be positively but weakly correlated. The evidence in Permanyer (2010) also suggests that socioeconomic polarization indices could be better predictors of the occurrence of conflicts than purely social polarization indices.

Finally, Esteban et al. (2012a) (see also Esteban et al. 2012b) assessed the empirical role of different notions of group division and group dispersion in explaining different types of conflict. Types of conflict differ according to whether it is the allocation of public or of private goods that is the object of dispute; their outcomes also vary according to levels of group cohesion. The exercise helps explore the empirical validity of the theoretical model described at the beginning of this chapter (see section 2). Esteban et al. (2012a) found that the extent of social polarization and fractionalization always empirically matters regardless of model specification. This suggests that sources of disputes both over public and over private goods matter in explaining conflict. Furthermore, and as predicted by the theoretical model, it is found that the greater the degree of relative publicness of conflict prizes, the greater the importance of polarization (which takes into account across-groups distances in utilities of public goods) relative to fractionalization (which does not).

5.10. CONCLUSION

Interest in the understanding and measurement of polarization is relatively recent. This chapter has reviewed some of the conceptual foundations of polarization; it has described how polarization can be measured in an economic, social, and/or hybrid socioeconomic perspective; it has reviewed briefly the empirical and theoretical polarization/conflict polarization literature; and it has distinguished polarization from inequality and other ways of considering distances between individuals and groups.

Given the increasing interest in the causes and effects of polarization, one would expect further advances in its understanding. On the measurement side, one can envisage, for instance, the development of techniques for decomposing polarization across socioeconomic groups, components of welfare, and features of the welfare distribution (such as modal structure, inequality, and measures of overlap); for making robust comparisons of polarization across time and societies; and for performing statistical inference on the levels and rankings of polarization. One can also foresee advances in the understanding of polarization levels and polarization dynamics, regarding, for instance, the role and the importance of socioeconomic changes (such as the effect of group sizes and distances) and the differential impact of alienation and identification across time. The dynamic impact of polarization on tensions and conflicts as well as the reverse impact are also plausible and important features of future research on polarization. Finally, the possible effect of growth, policy, and redistribution (in particular) on polarization, as opposed to other aspects of the income distribution, would also seem to constitute important areas for future research on polarization.

ACKNOWLEDGMENTS

The authors are grateful to Tony Atkinson, François Bourguignon, Satya Chakravarty, Joan Esteban, and Ignacio Permanyer Ugartemendia for helpful comments. This work was carried out with support from SSHRC and FRQSC.

REFERENCES

Acemoglu, D., Autor, D., 2012. What does human capital do? A review of Goldin and Katz's the race between education and technology. NBER Working Papers 17820, National Bureau of Economic Research, Inc.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., Wacziarg, R., 2003. Fractionalization. J. Econ. Growth 8, 155–194.

Amiel, Y., Cowell, F., 1992. Measurement of income inequality: experimental test by questionnaire. J. Public Econ. 47, 3–26.
Anderson, G., 2010. Polarization of the poor: multivariate relative poverty measurement sans frontiers. Rev. Income Wealth 56, 84–101.
Anderson, G., 2011. Polarization measurement and inference in many dimensions when subgroups cannot be identified. Economics: The Open-Access, Open-Assessment E-Journal. 5. http://dx.doi.org/10.5018/economics-ejournal.ja.2011-11.
Apouey, B., 2007. Measuring health polarization with self-assessed health data. Health Econ. 16, 875–894.
Aristotle-350. Politics, vol. 4, Part XI.
Atkinson, A., 1970. On the measurement of inequality. J. Econ. Theory 2, 244–263.
Autor, D.H., Katz, L.F., Kearney, M.S., 2008. Trends in U.S. wage inequality: revising the revisionists. Rev. Econ. Stat. 90, 300–323.
Blackburn, M., Bloom, D., 1995. What is happening to the middle class? Am. Demograph. 7, 19–25.
Chakravarty, S.R., 2009. In: Inequality, Polarization and Poverty. In: Advances in Distributional Analysis: Economic Studies in Inequality, Social Exclusion and Well-Being, vol. 6. Springer, New York.
Chakravarty, S.R., D'Ambrosio, C., 2010. Polarization ordering of income distributions. Rev. Income Wealth 56, 47–64.
Chakravarty, S., Maharaj, B., 2011a. Subgroup decomposable inequality indices and reduced-form indices of polarization. Keio Econ. Stud. 47, 57–83.
Chakravarty, S.R., Maharaj, B., 2011b. Measuring ethnic polarization. Soc. Choice Welfare 37, 431–452.
Chakravarty, S.R., Maharaj, B., 2012. Ethnic polarization orderings and indices. J. Econ. Interact. Coord. 7, 99–123.
Chakravarty, S., Maharaj, B., 2013. Generalized Gini polarization indices for an ordinal dimension of human well-being. In: Contemporary Issues and Ideas in Social Sciences, 6th Economic Theory and Policy Conference.
Chakravarty, S.R., Majumder, A., 2001. Inequality, polarization and welfare: theory and applications. Aust. Econ. Pap. 40, 1–13.
Chakravarty, S.R., Majumder, A., Roy, S., 2007. A treatment of absolute indices of polarization. Jap. Econ. Rev. 58, 273–293.
Chakravarty, S.R., Chattopadhyay, N., Maharaj, B., 2010. Inequality and polarization: an axiomatic approach. In: Deutsch, J., Silber, J. (Eds.), In: The Measurement of Individual Well-Being and Group Inequalities: Essays in Memory of Z. M. Berrebi, Routledge, London (Chapter 4).
Collier, P., Hoeffler, A., 2004. Greed and grievance in civil war. Oxf. Econ. Pap. 56, 563–595.
Colomer, J.H., 2001. Political Institutions, Democracy and Social Choice. Comparative Politics. Oxford University Press, USA.
Creamer, R., 2007. Listen to Your Mother; Stand Up Straight: How Progressives Can Win. Seven Locks Press, Santa Ana, CA.
Deutsch, J., Silber, J., 2010. Analyzing the impact of income sources on changes in bi-polarization. In: Deutsch, J., Silber, J. (Eds.), In: The Measurement of Individual Well-Being and Group Inequalities: Essays in Memory of Z. M. Berrebi, Routledge, London (Chapter 6).
Donaldson, D., Weymark, J., 1980. A single-parameter generalization of the Gini indices of inequality. J. Econ. Theory 22, 67–86.
Duclos, J.-Y., Araar, A., 2006. Poverty and Equity: Measurement, Policy, and Estimation with DAD. Springer and IDRC, Berlin and Ottawa.
Duclos, J.-Y., Échevin, D., 2005. Bi-polarization comparison. Econ. Lett. 87, 249–258.
Duclos, J.-Y., Esteban, J.-M., Ray, D., 2004. Polarization: concepts, measurement, estimation. Econometrica 72, 1737–1772.
Easterly, W., 2001. The middle class consensus and economic development. J. Econ. Growth 6, 317–335.
Easterly, W., Levine, R., 1997. Africa's growth tragedy: policies and ethnic divisions. Q. J. Econ. 112, 1203–1250.
Esteban, J., Ray, D., 2011a. A model of ethnic conflict. J. Eur. Econ. Assoc. 9, 496–521.
Esteban, J., Ray, D., 2012. Comparing Polarization Measures. Oxford University Press, Oxford.
Esteban, J.-M., Ray, D., 1994. On the measurement of polarization. Econometrica 62, 819–851.
Esteban, J.-M., Ray, D., 1999. Conflict and distribution. J. Econ. Theory 87, 379–415.
Esteban, J.-M., Ray, D., 2008. Polarization, fractionalization and conflict. J. Peace Res. 45, 163–182.

Esteban, J.-M., Ray, D., 2011b. Linking conflict to inequality and polarization. Am. Econ. Rev. 101, 1345–1374.

Esteban, J.-M., Gradín, C., Ray, D., 2007. An extension of a measure of polarization, with an application to the income distribution of five OECD countries. J. Econ. Inequal. 5, 1–19.

Esteban, J., Mayoral, L., Ray, D., 2012a. Ethnicity and conflict: an empirical study. Am. Econ. Rev. 102, 1310–1342.

Esteban, J., Mayoral, L., Ray, D., 2012b. Ethnicity and conflict: theory and facts. Science 336, 858–865.

Foster, J.E., Wolfson, M.C., 2010/1992. Polarization and the decline of the middle class: Canada and the U.S. J. Econ. Inequal. 8, 247–273.

Gigliarano, C., Mosler, K., 2009. Constructing indices of multivariate polarization. J. Econ. Inequal. 7, 435–460.

Glaeser, E.L., Ward, B.A., 2006. Myths and realities of American political geography. J. Econ. Perspect. 20, 119–144.

Gradín, C., 2000. Polarization by subpopulation in Spain: 1973-91. Rev. Income Wealth 46, 457–474.

Harsanyi, J.C., 1953. Cardinality utility in welfare economics and in the theory of risk-taking. J. Polit. Econ. 61, 434–435.

Herfindahl, O.C., 1950. Concentration in the U.S. steel industry. Ph.D. thesis, Columbia University.

Hirschman, A.O., 1980. National Power and the Structure of Foreign Trade. California Library reprint series. University of California Press, Berkeley, CA.

Horowitz, D.L., 1985. Ethnic Groups in Conflict. University of California Press, Berkeley, CA.

Kharas, H., 2010. The emerging middle class in developing countries. OECD Development Centre Working Papers, OECD Development Centre.

Kolm, S.-C., 1969. The optimal production of justice. In: Margolis, J., Guitton, S. (Eds.), Public Economics: An Analysis of Public Production and Consumption and Their Relations to the Private Sectors. MacMillan, London, pp. 145–200.

Krstcha, M., 1994. A new compromise measure of inequality. In: Eichhorn, W. (Ed.), Models and Measurement of Welfare and Inequality. Springer, Heidelberg.

Lasso de la Vega, C., Urrutia, A.M., 2006. An alternative formulation of the Esteban-Gradín-Ray extended measure of polarization. J. Income Distrib. 15, 42–56.

Lasso de la Vega, C., Urrutia, A.M., Henar, D., 2010. Unit consistency and bipolarization of income distributions. Rev. Income Wealth 56, 65–83.

Leckie, N., 1988. The Declining Middle and Technological Change: Trends in the Distribution of Employment Income in Canada, 1971-84. Economic Council of Canada, Ottawa.

Levy, F., 1987. The middle class: is it really vanishing? Brookings Rev. 5, 17–21.

Love, R., Wolfson, M.C., 1976. Income Inequality: Statistical Methodology and Canadian Illustrations. Catalogue 13559 Occasional, Statistics Canada, Ottawa, Canada.

Makdissi, P., Mussard, S., 2010. Rank-dependent measures of bi-polarization and marginal tax reforms. In: Deutsch, J., Silber, J. (Eds.), The Measurement of Individual Well-Being and Group Inequalities: Essays in Memory of Z. M. Berrebi. Routledge, London (Chapter 5).

Mauro, P., 1995. Corruption and growth. Q. J. Econ. 110, 681–712.

McCarty, N., Poole, K.T., Rosenthal, H., 2006. Polarized America: The Dance of Ideology and Unequal Riches. The MIT Press, Cambridge, MA.

Mishel, L., Schmitt, J., Shierholz, H., 2013. Assessing the job polarization explanation of growing wage inequality. Working Papers, Economic Policy Institute.

Montalvo, J.G., Reynal-Querol, M., 2002. Why ethnic fractionalization? Polarization, ethnic conflict and growth. Economics Working Paper 660, Universitat Pompeu, Fabra.

Montalvo, J.G., Reynal-Querol, M., 2003. Religious polarization and economic development. Econ. Lett. 80, 201–210.

Montalvo, J.G., Reynal-Querol, M., 2005a. Ethnic polarization, potential conflict, and civils wars. Am. Econ. Rev. 95, 796–816.

Montalvo, J.G., Reynal-Querol, M., 2005b. Ethnic diversity and economic development. J. Dev. Econ. 76, 293–323.

Montalvo, J.G., Reynal-Querol, M., 2008. Discrete polarization with an application to the determinants of genocides. Econ. J. 118, 1835–1865.

Morris, M., Bernhardt, A., Handcock, M., 1994. Economic inequality: new methods for new trends. Am. Sociol. Rev. 59, 205–219.

Nissanov, Z., Poggi, Z., Silber, J., 2010. Measuring bi-polarization and polarization: a survey. In: Deutsch, J., Silber, J. (Eds.), The Measurement of Individual Well-Being and Group Inequalities: Essays in Memory of Z. M. Berrebi. Routledge, London (Chapter 3).
Omalley, J., Blunden, A., 1970. Works of Karl Marx 1843: Critique of Hegel's Philosophy of Right. Cambridge University Press, Cambridge.
Permanyer, I., 2010. The conceptualization and measurement of social polarization. J. Econ. Inequal. 10, 45–74.
Permanyer, I., D'Ambrosio, C., 2013. Measuring social polarization with ordinal and categorical data. J. Public Econ. Theory. http://dx.doi.org/10.1111/jpet.12093.
Porta, R.L., De Silanes Shleifer, F.L., Vishny, R., 1999. The quality of government. J. Law Econ. Org. 15, 222–279.
Reynal-Querol, M., 2002. Ethnicity, political systems and civil wars. J. Conflict Resolut. 46, 29–54.
Rodríguez, J.G., Salas, R., 2003. Extended bi-polarization and inequality measures. In: Inequality, Welfare and Poverty: Theory and Measurement, vol. 9, pp. 69–83.
Sen, A.K., 1973. On Economic Inequality. Clarendon Press, Oxford.
Thurow, L.C., 1984. The disappearance of the middle class. New York Times, 2, February 5.
Wang, Y.-Q., Tsui, K.-Y., 2000. Polarization ordering and new classes of polarization indices. J. Public Econ. Theory 2, 349–363.
Wolfson, M.C., 1994. When inequalities diverge. Am. Econ. Rev. 84, 353–358.
Wolfson, M.C., 1997. Divergent inequalities: theory and empirical results. Rev. Income Wealth 43, 401–421.
Yitzhaki, S., 1983. On an extension of the Gini inequality index. Int. Econ. Rev. 24, 617–628.
Zhang, X., Kanbur, R., 2001. What difference do polarization measures make? An application to china. J. Dev. Stud. 37, 85–98.
Zheng, B., 2007. Unit-consistent decomposable inequality measures. Economica 74, 97–111.

CHAPTER 6

Statistical Methods for Distributional Analysis

Frank A. Cowell*, Emmanuel Flachaire†
*STICERD, London School of Economics, London, UK
†Aix-Marseille Université, AMSE & IUF, Marseille, France

Contents

Handbook of Income Distribution, Volume 2A
ISSN 1574-0056, http://dx.doi.org/10.1016/B978-0-444-59428-0.00007-2

Abstract

This chapter discusses the formal and informal techniques that are commonly used to give quantitative answers in the field of distributional analysis. To this end, it covers subjects including inequality, poverty, and the modeling of income distributions. It also deals with parametric and nonparametric approaches and the way in which imperfections in data may be handled in practice.

Keywords

Goodness of fit, Parametric modeling, Nonparametric methods, Dominance criteria, Welfare indices, Inequality measure, Poverty measure, Influence function, Hypothesis testing, Confidence intervals, Bootstrap

JEL Classification Codes

D31, D63, C10

6.1. INTRODUCTION

This chapter is about the techniques, formal and informal, that are commonly used to give quantitative answers in the field of distributional analysis—covering subjects such as inequality, poverty, and the modeling of income distributions.

At first sight, this might not appear to be the most exciting of topics. Discussing statistical and econometric techniques could appear to be purely secondary to the important questions in distributional analysis. However, this is not so. At a very basic level, without data what could be done? Clearly if there were a complete dearth of quantitative information about income and wealth distributions, we could still talk about inequality, poverty, and principles of economic justice. But theories of inequality and of social welfare would stay as theories without practical content. Knowing how to use empirical evidence in an appropriate manner is essential to the discourse about the welfare economics of income distribution and to the formulation of policy. Furthermore, understanding the nature and the limitations of the data that are available—or that may become available—may help to shape one's understanding of quite deep points about economic inequality and related topics; good practice in quantitative analysis can foster the development of good theory.

6.1.1 Why Statistical Methods?

If we carry out a simple computation of the values of an inequality or poverty measure, computed from two different samples, we will usually find greater inequality or poverty in one sample, *even if the two samples come from the same population*. Clearly simple computation alone is not enough to draw useful conclusions from the raw data: Statistical methods are required to test the hypothesis that the two values are not statistically different. For instance, Table 6.1 reports the values of the Gini and Theil inequality indices,[1] with confidence intervals at 95%, computed from two samples of 1000 observations drawn from the same distribution: The two samples are independent, with observations drawn independently from the Singh–Maddala distribution with parameters $a=2.8$, $b=0.193$, and $q=1.7$, which closely mimics the net income of German households, up to a scale factor (Brachmann et al., 1996). Clearly the values of the Gini and Theil indices are greater in sample 1 than in sample 2. However, the confidence intervals (in brackets) intersect for both inequality measures, which leads us to not reject the hypothesis that the level of inequality is the same in the two samples.

There is a wide variety of inequality indices in common use. Different indices, with different properties, could lead to opposite conclusions in practice. Lorenz curves comparisons can be very useful because a (relative) Lorenz curve always lying above another one implies that any comparisons of relative inequality measures would lead to similar conclusions—a result that holds for any inequality measures respecting anonymity, scale invariance, replication invariance, and the transfer principle (Atkinson, 1970). In practice, we have on hand a finite number of observations, and empirical Lorenz dominance can be observed many times when the two samples come *from the same population*. In the case of two independent samples of 1000 observations drawn from the same Singh–Maddala distribution, we obtain sample Lorenz dominance 22% of cases. Dardanoni and Forcina (1999) argued that it can be as high as 50% of cases due to the fact that empirical Lorenz curve ordinates are typically strongly correlated. This demonstrates the need to use statistical methods.

The point about simple computation being insufficient is also easily demonstrated in the case of Lorenz curves: Figure 6.1 shows the difference between the empirical Lorenz curves obtained from two independent samples drawn from the same distribution, with confidence intervals at 95% calculated at the population proportions $q=0.01, 0.02, \ldots, 0.99$. The ordinates are always positive, so it is clear that one empirical Lorenz curve always dominates the other. However, the confidence intervals show that *each* Lorenz

Table 6.1 Inequality indices with confidence intervals at 95%

	Sample 1	**Sample 2**
Gini	0.303	0.285
	[0.286; 0.320]	*[0.271; 0.299]*
Theil	0.158	0.135
	[0.133; 0.183]	*[0.120; 0.151]*

[1] For formal definitions see Equations (6.51), (6.69), and (6.70).

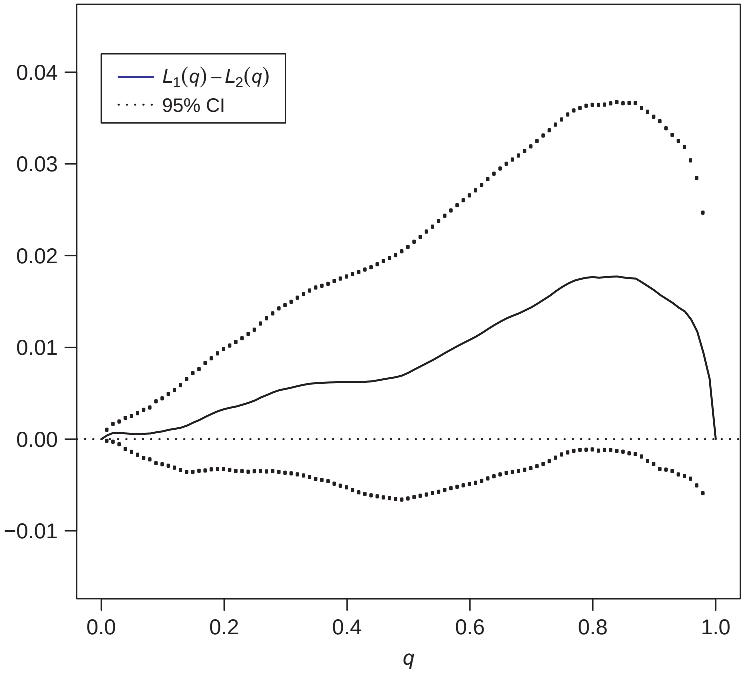


Figure 6.1 Difference between two empirical Lorenz curves, $\hat{L}_1(q) - \hat{L}_2(q)$, with 95% confidence intervals. The samples are drawn from the same distribution.

curve ordinate difference is never significantly different from zero; as a result Lorenz dominance in the population is not as clear as simple computation from the sample might suggest. To be able to make conclusions on dominance or nondominance, we need to test simultaneously that *all* ordinate differences are statistically greater than zero, or not less than zero. Appropriate test statistics need to be used to make such multiple comparisons.

In this chapter, we will provide a survey of the theory and methods underlying good practice in the statistical analysis of income distribution. We also offer a guide to the tools that are available to the practitioner in this field.

6.1.2 Basic Notation and Terminology

Throughout the chapter, certain concepts are used repeatedly, and so it is convenient to list some of the terms that are used repeatedly.

- *Income* y. Here "income"; this is merely a convenient shorthand for what in reality may be earnings, wealth, consumption, or something else. We will suppose that y belongs to a set $\mathbb{Y} = [\underline{y}, \overline{y})$, an interval on the real line $\mathbb{R}$.
- *Population proportion* q. For convenience, we will write $q \in \mathbb{Q} := [0,1]$.
- *Distribution* F. This is the cumulative distribution function (CDF) so that, for any $y \in \mathbb{Y}$, $F(y)$ denotes the proportion of the population that has income y or less. Where

the density is defined, we will write the density at $y \in \mathbb{Y}$ as $f(y)$. The set of all distribution functions will be denoted $\mathbb{F}$.

- *Indicator function* $\iota(\cdot)$. Suppose there is some logical condition D, which may or may not be true. Then $\iota(\cdot)$ is defined as:

$$\iota(D) = \begin{cases} 1 & \text{if } D \text{ is true} \\ 0 & \text{if } D \text{ is not true} \end{cases} \tag{6.1}$$

6.1.3 A Guide to the Chapter

We begin with a discussion of some of the general data issues that researchers should bear in mind (Section 6.2). Section 6.3 deals with the issues that arise if we want to try to "model" an income distribution: The motivation for this is that sometimes it makes sense to approach the analysis of income distributions in two stages: (1) using a specific functional form or other mathematical technique to capture the evidence about the income distribution in an explicit model, and (2) making inequality comparisons in terms of the modeled distributions. Section 6.4 deals with the general class of problem touched on in our little example outlined in Table 6.1: The emphasis is on hypothesis testing using sample data, and we cover both inequality and poverty indices. As a complement to this, Section 6.5 deals with the class of problem highlighted in Figure 6.1: We look at a number of "dominance" questions that have a similarity with the Lorenz problem described there. Section 6.6 returns to mainly data-related questions: how one may deal with some of the practical issues relating to imperfections in data sets. Finally, in Section 6.7, we draw together some of the main themes that emerge from our survey of the field.

6.2. DATA

6.2.1 Data Sources

It is not difficult to imagine instances where there is a known, finite set of income-receivers and where the income associated with each income-receiver is observable (example: if there are 50 states in a federation and one wishes to analyze the change in the distribution of announced military contracts awarded by the federal government among those 50 states). Under those circumstances, a complete enumeration of the relevant "population" (the 50 states) is possible, and the income of each member of this "population" is measured with complete accuracy (from the federal government announcement). There is very little to do in terms of statistical analysis and no data problem. But this kind of example is rarely encountered in practice and might be dismissed as somewhat contrived. It is much more common to have to deal with situations in which an enumeration of the population is impossible and we have to rely on some sort of sample.

6.2.1.1 Administrative Data

Governments and government agencies have long published summaries of income distributions in grouped form; in many countries, official data providers have gone further and made available to researchers microdata from official sources that could be used, for example, to analyze the distribution of income and wealth. The data made available in this way used to be of similar size to sample surveys (discussed later). However, it is increasingly the case that very large data sets have been opened up for research, an order of magnitude larger—effectively complete collections of administrative data rather than official samples from them. It might be tempting to treat these as methodologically equivalent to the complete enumeration case described earlier. But this would overlook two points. First, administrative data will only contain what is legally permissible and what government agencies find convenient to release. If, for example, one is interested in the distribution of personal incomes, a very large data set of tax records could be extremely useful, but it will miss many of those persons who are not required to file tax returns. Second, the design of the data set may not match what the social scientist or economist would wish: For example, if one wishes to adjust the data to allow for differences in need according to the type of household or family in which each person lives, the required information for constructing an appropriate equivalence scale may not be present in the same data set.

6.2.1.2 Survey Data

The problems from administrative data stem largely from the fact that the data are the by-product of information gathered for other purposes. It is clear that specially designed surveys have a potential advantage in this respect. However, although surveys are usually purpose-built (and often purpose-built using advice from social scientists), one also has to be cautious about their limitations. This concerns the smaller size and worse response rate than the administrative data counterparts. Once again the survey design may exclude some sections of the population (a survey based on households would obviously miss people who are homeless and those in institutions), and where there is an attempt to create longer series of surveys, the criteria for the design of contemporary surveys may follow a standardized format determined by conventions that are no longer relevant.

6.2.2 Data Structure

In implementing the statistical criteria discussed in this chapter, one needs to be clear about the relevant assumptions concerning the way the sample was drawn.

6.2.2.1 Simple Design

In the majority of this chapter, we will take it that simple random sampling is an appropriate assumption. By this we mean that the sample has been designed in such a way that each member of the population has an equal probability of being included in the sample.

This can be taken as an ideal case that enables one to focus on the central issues of statistical inference. Even the supposedly "ideal" case may not be ideal in practice if the sampling frame is inappropriate—it could be out of date or it could be specified in such a way that part of the population is excluded (see the remarks earlier about homeless people).

6.2.2.2 Complex Design

In practice, there are often simple practical reasons why something other than simple random sampling is used.[2] Two features in particular are often built into the design of the sample. *Clustering* the observations by geographical location may reduce the costs of running the survey, both in terms of initial visits to carry out the survey and in follow-up visits for monitoring and completing missing information. *Stratification* is a common technique for deliberately oversampling certain categories of respondent to ensure that there is adequate representation in the combined sample of certain types of individuals or households that are of particular interest but that are likely to show up comparatively rarely either because they are genuinely rare in the population or because they are less likely to respond to the survey (for example, it is commonly found that richer households are overrepresented in the "nonresponse" category, and if one were just to ignore that possibility, there would be the danger of having a biased sample). In effect one divides up the population of interest into subpopulations and chooses a sample for each subpopulation—each stratum—at an appropriate sampling rate.

Although the assumption of a simple random sample sweeps aside practical problems associated with the design of the survey, this idealized case gives us a good base for explaining the core issues in estimation and inference. At appropriate points in Sections 6.4 and 6.5, we will comment on the extensions to the complex data case and other related issues.[3]

Other problems with the data merit special discussion. We briefly outline the nature of these problems here and then return to a formal analysis of them (in Section 6.6) after we have extensively discussed conventional inference problems in the preceding sections.

6.2.3 Data Problems

6.2.3.1 Measurement Error and Data Contamination

Measurement error in income distribution analysis can be handled in a way similar to measurement error in other contexts. Observed income is true income adjusted by an error term (Chesher and Schluter, 2002), and the resulting model resembles the problem of

[2] See Deaton (1997) for a full discussion of the issues involved.

[3] For example, if the data are based on a simple survey of households, but one wants to infer something about the distribution of individuals, one needs to weight each household observation by an amount proportional to the number of persons in the household; this structure is similar to the weights introduced by stratification.

decomposition by factor source; *data contamination* can be represented as a mixture of a true distribution and a contamination distribution. The resulting model resembles the problem of decomposition by population subgroup (Cowell, 2000; Cowell and Fiorio, 2011). However, the appropriate model for analyzing this second type of problem uses tools that are useful for the analysis of questions beyond the narrow data contamination question. This will be discussed in Sections 6.4–6.6.

6.2.3.2 Incomplete Information

In many practical applications, we need to deal with situations in which some parts of the sample space are excluded completely from the sample data or where information in part of the sample is missing; for convenience we will refer to this part of the sample as the "excluded" subset, even though some information may be available. The exclusion of information may be imposed by the data provider, for example, because of reasons of confidentiality, or it may be introduced by the researcher to deal pragmatically with some other problem in the data.

Table 6.2, taken from Cowell and Victoria-Feser (2003), sets out the main cases that are of interest. There are two principal issues for the researcher to consider, as follows:

Boundaries of the excluded subset. What determines the boundaries of the excluded subset of the sample space? There are two possible cases summarized in the rows of Table 6.2: (i) a subset of $\mathbb{Y}$ is specified, or (ii) a subset of $\mathbb{Q}$ is specified. In the first case, the income-boundaries of the excluded subset $(\underline{z}, \overline{z})$ are fixed, but the proportions of the excluded subsets $(\underline{\beta}, \overline{\beta})$ are unknown, although these proportions can be estimated if enough information is available. In the second case, the boundaries of the excluded sample are fixed by the trimming proportions in the lower and upper tail $(\underline{\beta}, \overline{\beta})$, and the incomes at the boundary of the excluded samples $(\underline{z}, \overline{z})$ are unknown.

Information in the excluded subset. There are several assumptions about the availability of information in the excluded part of the sample. The situation is going to depend on the particular problem in hand, and the principal cases are summarized in the columns of Table 6.2. At one extreme, the excluded subset is just *terra incognita* (left-hand column). On the other hand, it may be that the data provider makes available several summary statistics related to excluded subset (right-hand column).

Table 6.2 Types of incomplete information

	Information about excluded sample		
	None	**Sample proportion**	**Multiple statistics**
Incomes below $\underline{z}$ and above $\overline{z}$ excluded	**A**	**B**	**C**
Lowest $100.\underline{\beta}\%$ and highest $100.\overline{\beta}\%$ excluded	**D**	**(E)**	**(F)**

So, in principle there are altogether six possible cases, but in practice only four are relevant:[4]

- Case A is the standard form of truncation.
- Case B represents "censoring"; in this case there are point masses at the boundaries $\left(\underset{\sim}{z}, \overline{z}\right)$ that estimate the population-share of the excluded part.[5]
- Case C is an extension of standard estimation problem with grouped data (Gastwirth et al., 1986).
- Case D represents the case of trimming.

The implications of these issues for distributional analysis are considered in Section 6.6.2.

6.2.4 Grouped Data

For reasons of economy and convenience, it used to be common practice for statistical offices to make income distribution data available only in grouped form (see Case C in Table 6.2). Typically, this would involve a simple table with a comprehensive set of pre-set income intervals, the numbers of individuals or households falling into each interval, and (sometimes) the average income associated with each interval. Tabulated data are less usual today, although researchers are increasingly using historical data to construct long-run time series. So, it is useful to consider briefly the analytical issues that arise in connection with this type of data.

One way of using such data effectively is to estimate the underlying income distribution using parametric modeling. This can be done either by using interpolation methods in each of the intervals (see, for example, Cowell, 2011) or by fitting a distribution to the bulk of the data—suitable parametric methods are discussed in Section 6.3.1. Nonparametric methods are necessarily quite limited because of the restrictions imposed by the data.[6] However, an interesting problem presented by any sort of grouped data is to compute bounds on inequality indices. One uses the available information to compute a maximum-inequality distribution and a minimum inequality distribution by making alternative extreme assumptions about the way the data are distributed within each interval (Cowell, 1991; Gastwirth, 1972, 1975).

[4] If, as is usual, trimming is something that is done voluntarily by the user rather than being imposed by the data provider, then Cases E and F are not relevant in practice.

[5] A standard example of this "top-coding" of some income components in the Current Population Survey—observations above a given value $\overline{Z}$ are recorded as $\overline{Z}$ (Polivka, 1998). In practical applications of distributional analysis, such as inferring inequality trends, researchers have adopted a number of work rounds such as multiplying top-coded values by a given factor (Autor et al., 2008; Lemieux, 2006) or attempting imputations for missing data (Burkhauser et al., 2010; Jenkins et al., 2011).

[6] The problem of statistical inference with grouped data is discussed in Hajargasht et al. (2012).

6.3. DENSITY ESTIMATION

The analysis of a probability density function is a powerful tool to describe several properties of a variable of interest. For instance, Figure 6.2 shows the estimated density function of GDP per capita in 121 countries across the world in 1988.[7] We can see that the density function is bimodal. The existence of two modes suggests that there are two distinct groups: one composed of the "richest" countries and another consisting of the "poorest." The second mode is much less pronounced than the first, which indicates that the two groups are not of the same size; there are relatively few "rich" countries and distinctly more "poor" countries. Further, the first mode is located just to the left of the value 0.5 on the *X*-axis, whereas the second is found at around 3. We can thus conclude from this figure that, on average in 1988, "rich" countries enjoyed a level of GDP per capita that was around three times the average, whereas that of "poor" countries was only half of the average level. It is clear from this example that much more information is

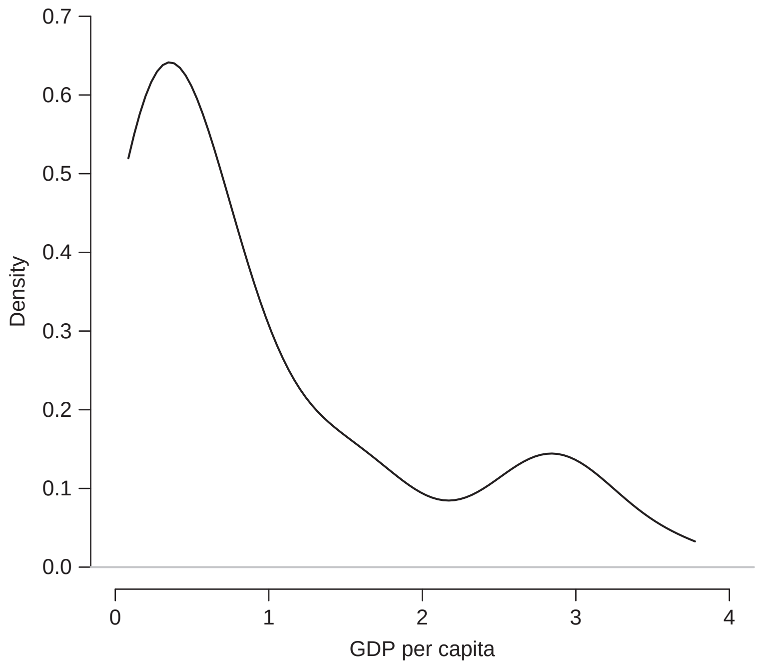


Figure 6.2 Kernel density estimation of GDP per capita in 121 countries normalized by the global mean.

[7] The data are unweighted (each country as equal weight) and are taken from the *Penn World Table* of Summers and Heston (1991). The horizontal axis is the per capita GDP for each country normalized by the (unweighted) mean over all countries.

available from the full distribution of a variable than the restricted information provided by standard descriptive statistics, as the mean, variance, skewness, or kurtosis, which summarize each limited properties of the distribution on single values.

In the multivariate case, the conditional density function can provide useful insights on the relationship between several variables. For instance, Figure 6.3 shows the estimated density functions of wages conditional to experience for individuals with the same level of education.[8] We can see that as experience increases, the conditional distribution becomes bimodal, and the gap between the two modes increases. It suggests that the population is composed by two subgroups, and the marginal impact of experience on wages is not the same for the two groups. A standard regression tracks the dynamics of the first moment of the conditional distribution only and then cannot highlight the features just described. Here, a linear regression of wages on experience would estimate the marginal impact of experience on the average of wages for all individuals, whereas the graphical analysis suggests that experience does not affect individual's wages identically. Mixture models of regressions would be more appropriate in such cases.

In practice, the functional form of the density function is often unknown and has to be estimated. For a long time, the main estimation method was mainly parametric. However, a parametric density estimation requires the choice of a functional form *a priori*, and most of them do not fit multimodal distribution. In the last two decades, nonparametric and semiparametric estimation methods have been extensively developed. They are often used in empirical studies now and allow us to relax the specific assumptions underlying parametric estimation method, but they require in general more data on hand.

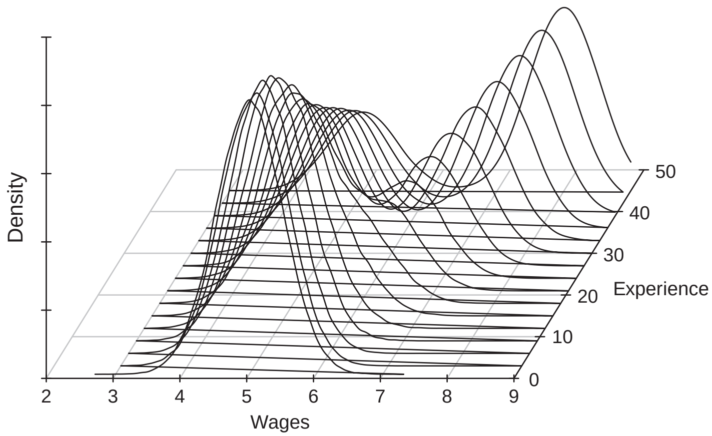


Figure 6.3 Conditional density estimates of wages on experience.

[8] The data are simulated.

In this section, we will present parametric, nonparametric, and semiparametric density estimation methods. Standard parametric methods are presented in Section 6.3.1, kernel density methods in Section 6.3.2, and finite mixture models in Section 6.3.3.

6.3.1 Parametric Estimation

We say that a random variable Y has a probability density function f, if the probability that Y falls between a and b is defined as:

$$\mathrm{Prob}(a < Y < b) = \int_a^b f(y)\mathrm{d}y,$$

where a and b are real values and $a < b$. The density function f is defined as nonnegative everywhere, and its integral over the entire space is equal to one.

A parametric estimation requires one to specify *a priori* a functional form of the density function, that is, to know the density function up to some parameters. The density function can then be written $f(y; \theta)$, where θ is a vector of k unknown parameters and y a vector of n observations. The problem remains of estimating θ; this is usually done by maximizing the likelihood to observe the actual values in a sample. If the data are independent and identically distributed (IID), the joint density function of n observations y_1, $y_2, \ldots, y_n$ is equal to the product of the individual densities:

$$f(y; \theta) = \prod_{i=1}^{n} f(y_i; \theta).$$

The estimation of the density function therefore requires the maximization of this function with respect to θ. Because the logarithmic function is positive and monotone, it is equivalent to maximize

$$\ell(y; \theta) = \log f(y; \theta) = \sum_{i=1}^{n} \log f(y_i; \theta),$$

from which the solution is often much simpler. This estimation method is known as the maximum likelihood method.

6.3.1.1 Pareto

Pareto (1895) initiated the modeling of income distribution with a probability density function[9] that is still in common use for modeling the upper tail of income and wealth distributions. Beyond a minimal income, he observed a linear relationship between the logarithm of the proportion of individuals with incomes above a given level and the

[9] What is now commonly described as *the* Pareto distribution is more precisely referred to as "Pareto type I." For other, more general, forms introduced by Pareto and their relationship to the Pareto type I see Cowell (2011) and Kleiber and Kotz (2003).

logarithm of this given level of income. This observation has been made in many situations and suggests a distribution that decays like a power function; such behavior characterizes a heavy-tailed distribution.[10] The Pareto CDF is given by

$$F(y;\alpha) = 1 - \left[\frac{y}{y_0}\right]^{-\alpha}, \quad y > y_0 \tag{6.2}$$

with density

$$f(y;\alpha) = \alpha y^{-\alpha-1} y_0^{\alpha} \tag{6.3}$$

If p_y denotes the proportion of the population with incomes greater than or equal to y (for $y \geq y_0$), then we have

$$\log p_y = \log A - \alpha \log y \tag{6.4}$$

where $A := y_0^{\alpha}$. The Pareto index α is the elasticity of a reduction in the number of income-receiving units when moving to a higher income class. The larger the Pareto index, the smaller the proportion of very high-income people. The Pareto distribution often fits wealth distributions and high levels of income well—see Figure 6.4—but it is not designed to fit low levels of income. Other distributions have then been proposed in the literature.

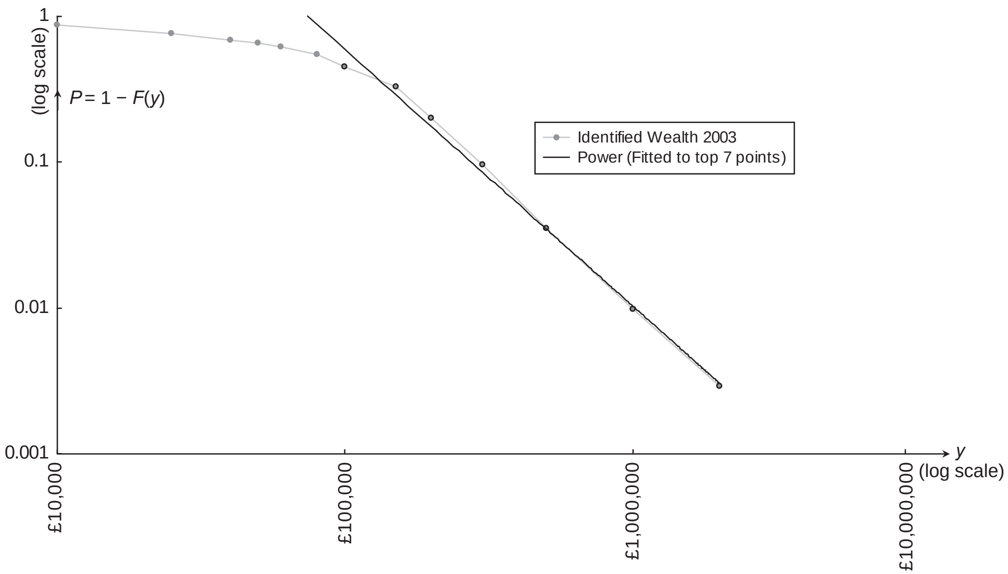


Figure 6.4 Pareto distribution. UK Identified wealth 2003. *Source: Inland Revenue Statistics 2006, Table 13.1*.

[10] A "heavy-tailed" distribution F is one for which $\lim_{y\to\infty} e^{\lambda y}[1 - F(y)] = \infty$, for all $\lambda > 0$: it has a tail that is heavier than the exponential.

6.3.1.2 Lognormal

Gibrat (1931) highlighted the central place of the lognormal distribution in many economic situations. His *law of proportionate effect* says that if the variation of a variable between two successive periods of time is a random proportion of the variable at the first period, then the variable follows a lognormal distribution.[11] He successfully fitted lognormal distributions with many different data sets, as for instance income, food expenditures, wages, bequests, rents, real estate, firm profits, firm size, family size, and city size. The lognormal distribution has then been very popular in empirical work[12] and is often appropriate for studies of wages—see Figure 6.5. However, the fit of the upper tail of more broadly based income distributions appears to be quite poor. The tail of the lognormal distribution decays faster than the Pareto distribution, at the rate of an exponential function rather than of a power function. It has led to the use of other distributions with two to five parameters to get a better fit of the data over the entire distribution.

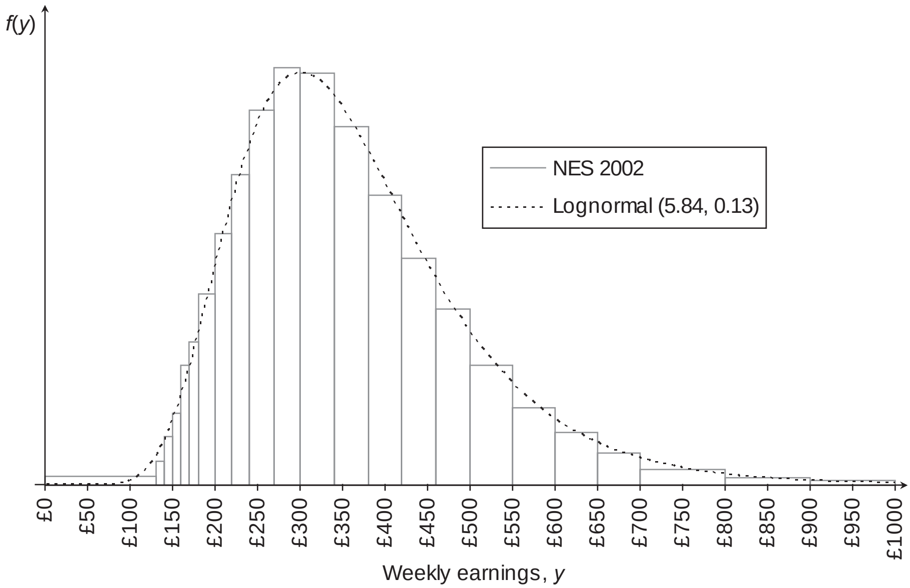


Figure 6.5 Lognormal distribution. UK Male Manual Workers on Full-Time Adult Rates. *Source: New Earnings Survey 2002, Part A Table A35.*

[11] If $X_t - X_{t-1} = \varepsilon_t X_{t-1}$, then $\sum_{i=1}^{n} \varepsilon_t = \sum_{i=1}^{n} \frac{X_t - X_{t-1}}{X_{t-1}} \approx \log X_n - \log X_0$. From a central limit theorem (CLT), log X_n follows asymptotically a Normal distribution.

[12] In addition, it has nice properties related to the measurement of inequality (Cowell, 2011), it is closely related to the Normal distribution, and it fits homogeneous subpopulations quite well (Aitchison and Brown, 1957).

6.3.1.3 Generalized Beta

The gamma and Weibull distributions have shown good fit in empirical studies.[13] The lognormal, gamma, and Weibull density functions are two-parameter distributions; they share the property that Lorenz curves do not intersect, contrary to what is observed in several datasets. To allow intersecting Lorenz curves, three-parameter distributions should be used, as the generalized gamma (GG), Singh–Maddala (SM), and Dagum distributions.[14] As shown by McDonald and Xu (1995), all the previously mentioned distributions are special or limiting cases of the five-parameter generalized beta distribution, defined by the following density function:

$$\mathrm{GB}(y; a, b, c, p, q) = \frac{|a| y^{ap-1}\left[1-(1-c)(y/b)^a\right]^{q-1}}{b^{ap} B(p, q)\left[1+c(y/b)^a\right]^{p+q}}, \tag{6.5}$$

for $0 < y^a < b^a/(1-c)$ and is equal to zero otherwise. $\mathrm{B}(p, y) := \int_0^1 t^{p-1}(1-t)^{q-1}\,\mathrm{d}t$ is the beta function, $0 \le c \le 1$ and b, p, and q are positive. Figure 6.6 shows graphically the relationships between distributions.[15] As an example of the paths through this diagram, take the case where $c=0$ in (6.5): we find the Generalized Beta of the first kind

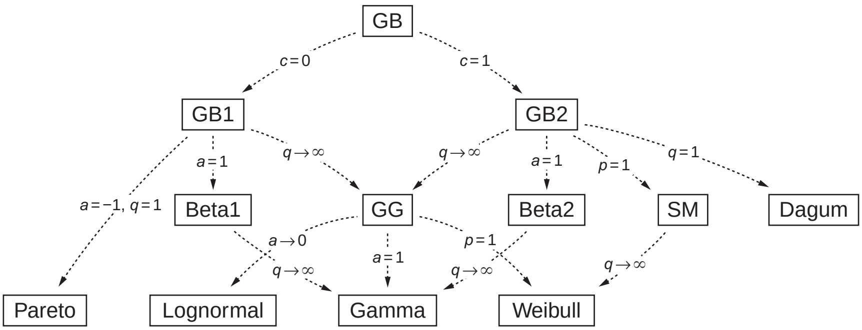


Figure 6.6 Parametric distributions tree. *Source: Bandourian et al., 2003.*

[13] Among others, Salem and Mount (1974) show that the gamma distribution fits better than the lognormal for income data in the United States for the years 1960 to 1969; Bandourian et al. (2003) found the Weibull distribution to be the best two-parameter distribution for income distribution in many countries

[14] See Stacy (1962), Singh and Maddala (1976), and Dagum (1977). The Singh–Maddala and Dagum distributions are also known as, respectively, the Burr 12 and Burr 3 distributions.

[15] GB1 and GB2 are, respectively, the generalized beta of the first and second kinds introduced by McDonald (1984). Beta1 and Beta2 are, respectively, the beta of first and second kinds. An alternative three-parameter approach that giving a good representation of income distributions in practice is provided by the Pareto–Lévy class (Dagsvik et al., 2013; Mandelbrot, 1960); unfortunately, except in a few cases, the probability distributions associated with this class cannot be represented in closed form.

$$\mathrm{GB1}(y; a, b, p, q) = \frac{|a| y^{ap-1}[1-(y/b)^a]^{q-1}}{b^{ap}\mathrm{B}(p, q)}. \tag{6.6}$$

Going a stage further, the special case of (6.6) with $a = 1$ gives the Beta distribution of the first kind

$$\mathrm{B1}(y; b, p, q) = \frac{y^{p-1}[1-(y/b)]^{q-1}}{b^{p}\mathrm{B}(p, q)}, \tag{6.7}$$

As an alternative route from (6.6), setting $a = -1$ and $q = 1$ we obtain $b^p y^{-p-1}$, which, with a change of notation, is clearly the density function of the Pareto type I distribution (6.3). For more details on continuous univariate distributions, see Johnson et al. (1994) and Kleiber and Kotz (2003).

Income distributions have been extensively estimated with parametric density functions in the literature; see, for instance, Singh and Maddala (1976), Dagum (1977, 1980, 1983), McDonald (1984), Butler and McDonald (1989), Majumder and Chakravarty (1990), McDonald and Xu (1995), Bantilan et al. (1995), Victoria-Feser (1995, 2000), Brachmann et al. (1996), Bordley et al. (1997), Tachibanaki et al. (1997), and Bandourian et al. (2003). In most of these empirical studies, the generalized beta of the second kind, the Singh–Maddala and the Dagum distributions perform better than other two/three-parameter distributions.

6.3.1.4 Goodness of Fit

Goodness-of-fit test statistics are used to test whether a given sample of data is drawn from an estimated probability distribution. They are used to know if an estimated density function is appropriate and fits the data well. Several statistics have been proposed in the literature. The well-known Pearson chi-squared statistic is defined as: $\chi^2 = \sum_{i=1}^{m} (O_i - E_i)^2 / E_i$, where O_i is the observed percentage in the ith histogram interval, E_i is the expected percentage in the ith histogram interval, and m is the number of histogram intervals. This measure summarizes discrepancies between frequencies given by a histogram obtained from the data and those expected from the estimated density function. A statistic not significantly different from zero suggests that the estimated density function fits well the unknown density function that generated the data. In finite samples, this statistic is known to have poor finite sample power properties, that is, to underreject when the estimated density function is not appropriate; see Stephens (1986). Then the Pearson chi-square test is usually not recommended as a goodness-of-fit test. Empirical distribution function (EDF)-based statistics perform better. Given a set of observations $\{y_1, y_2, \ldots, y_n\}$, the EDF is defined as

$$F^{(n)}(y) = \frac{1}{n} \sum_{i=1}^{n} \iota(y_i \le y) \tag{6.8}$$

where $\iota(\cdot)$ is the indicator function defined in (6.1). When the observations are independently and identically distributed, it is a consistent estimator of the CDF. EDF-based statistics measure the discrepancy between the EDF and the estimated CDF. They are not sensitive to the choice of the histogram's bins as in the Pearson chi-squared statistic. For instance, the Kolmogorov–Smirnov statistic is equal to

$$\sup_y |F^{(n)}(y) - F(y, \hat{\theta})| \tag{6.9}$$

where $F(y, \hat{\theta})$ is an estimated CDF from a parametric family function with parameters θ. Other statistics can be expressed as

$$n\int_{-\infty}^{\infty} \left(F^{(n)}(y) - F(y, \hat{\theta})\right)^2 w(y)\mathrm{d}F(y, \hat{\theta}), \tag{6.10}$$

where $w(y)$ is a weighting function. The Cramér–von Mises statistic corresponds to the special case $w(y) = 1$, while the Anderson–Darling statistic puts more weight in the tails, with $w(y) = \left[F(y, \hat{\theta})\left(1 - F(y, \hat{\theta})\right)\right]^{-1}$. In finite samples, the Anderson–Darling test statistic outperforms the Cramér–von Mises statistic, which in turn outperforms the Kolmogorov–Smirnov test statistic, see Stephens (1986).

In distributional analysis, goodness-of-fit test statistics and inequality measures do not usually share the same intellectual foundation. The former are based on purely statistical criteria, whereas the latter are typically based on an axiomatic that may be associated with social welfare analysis or other formal representations of inequality in the abstract. Cowell et al. (2014) developed a family of goodness-of-fit tests founded on standard tools from the economic analysis of income distributions, defined as:

$$G_\xi = \frac{1}{(\xi^2 - \xi)} \sum_{i=1}^{n} \left[\left[\frac{u_i}{\mu_u}\right]^\xi \left[\frac{2i}{n+1}\right]^{1-\xi} - 1 \right], \tag{6.11}$$

where $\xi \in \mathbb{R} \setminus \{0, 1\}$ is a parameter, $u_i = F(y_{(i)}; \hat{\theta}), \mu_u = \frac{1}{n}\sum_{i=1}^{n} u_i$ and $y_{(i)}$ is the ith order statistic (the ith smallest observation). G_ξ is closely related to Generalized Entropy (GE) inequality indices—see Equations (6.49)–(6.51). GE inequality measures are divergence measures between the EDF and the most equal distribution, where everybody gets the same income. They tell us how far an empirical distribution is from the most equal distribution.[16] Goodness-of-test statistics G_ξ are divergence measures between the EDF and an estimated parametric CDF. They tell us how far an empirical distribution is from an

[16] See Cowell et al. (2013) for an extension to the choice of any other "reference" distribution, giving for instance inequality measures telling us how far an empirical distribution is from the most *unequal* distribution.

estimated parametric distribution.[17] It has excellent size and power properties as compared with other, commonly used, goodness-of-fit tests. It has the further advantage that the profile of the G_ξ statistic as a function of ξ can provide valuable information about the nature of the departure from the target family of distributions, when that family is wrongly specified.

6.3.2 Kernel Method

6.3.2.1 From Histogram to Kernel Estimator

Histograms are the most widely used nonparametric density estimators. However, they have several drawbacks that kernel density method allows us to handle.

Figure 6.7 illustrates several problems arising with histograms, using GDP per capita in 121 countries in 1988 (solid line). The left plot is given with five bins of the same length between 0 and 5. The middle plot is similar but the position of the bins has changed; they are between –0.5 and 4.5. The two pictures are very different, even if they estimate the same distribution. The left panel shows a unimodal distribution, whereas the middle panel shows a bimodal distribution. Histograms are then sensitive to the point at which we start drawing bins. The right panel is given with 10 bins of the same length between 0 and 5. Once again, it gives a different picture of the same distribution. Histograms are thus sensitive to the number of bins used, which is also relatively arbitrary. Moreover, and most obviously, the pictures given by histograms provide discontinuities at the edge of each bin, which may not be an appropriate property of the true underlying distribution.

To avoid having to make an arbitrary choice on the position and the number of bins, we can use intervals that may overlap, rather than being separate from each other. The principle here is to estimate a density function at one point by counting the number of observations that are close to this evaluation point. For a sample of n observations, $y_1, \ldots, y_n$, the naive density estimator is given by:

$$\hat{f}(y) = \frac{1}{nh}\sum_{i=1}^{n} \iota\left(y - \frac{h}{2} < y_i < y + \frac{h}{2}\right), \tag{6.12}$$

where h is the width of the intervals and $\iota(\cdot)$ is the indicator function (6.1). In this equation, the estimate of the density at point y is given by the proportion of observations that

[17] The term in the first bracket in (6.11) is related to the CDF, $u_i/\mu_u = F(y_{(i)}, \hat{\theta})/\mu_u$, whereas the term in the second bracket is related to the EDF, $2i/(n+1) = v_i/\mu_v$ where $v_i = \hat{F}^{(n)}(y_{(i)}) = i/n$ and $\mu_v = n^{-1}\sum_{i=1}^{n} v_i = \sum_{i=1}^{n} i/n^2 = (n+1)/(2n)$. Using the most equal distribution and replacing u_i and v_i by their q-quantile counterparts, $F^{-1}(q_i = i/n, \hat{\theta}) = \hat{\mu}_y$ and $\hat{F}^{(n)^{-1}}(q_i = i/n) = y_{(i)}$, give GE measures. Note that u_i and v_i have bounded support ($u_i, v_i \in [0, 1]$), a property required to show that the asymptotic distribution of G_ξ exists, see Davidson (2012).

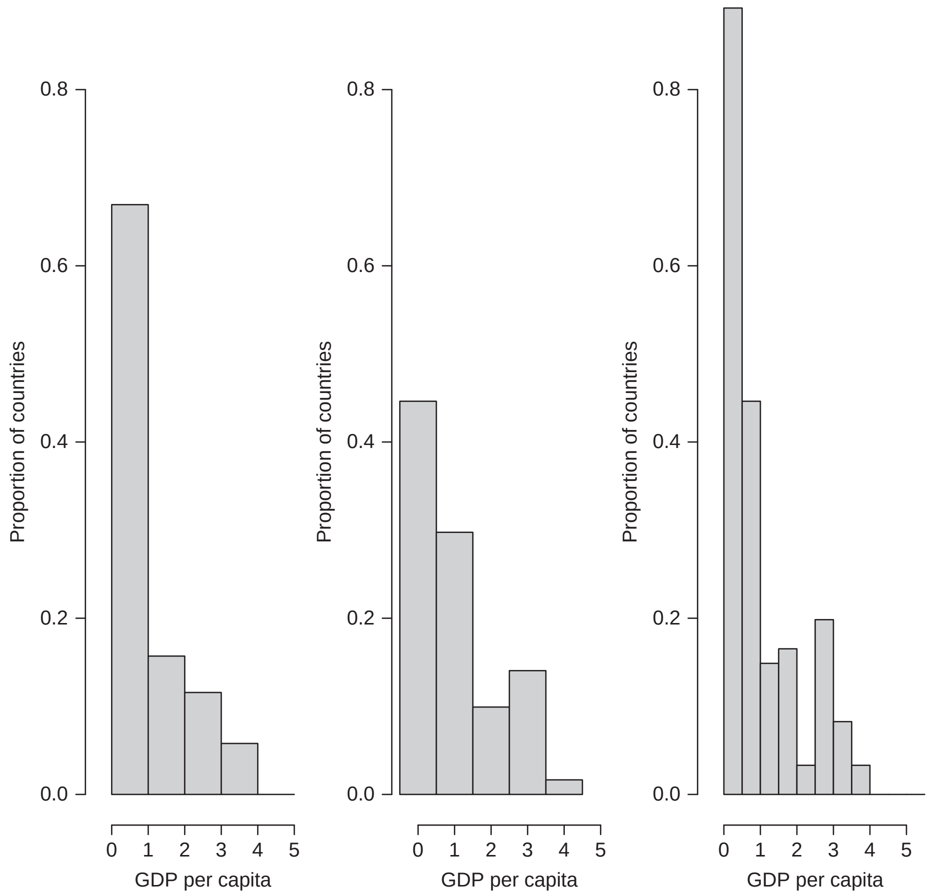


Figure 6.7 Histogram's sensitivity to the position and the number of bins.

are within a distance of $h/2$ or less from point y. The global density is obtained by sliding this window of width h along all the evaluation points.

Figure 6.8 presents the naive estimation of the density of GDP per capita across different countries in 1988. Compared to histograms, the naive estimator reveals much more detail about the curvature of the density function. However, discontinuities are still present.

The kernel estimator is a generalization of the naive estimator, which allows us to overcome the problem of differentiability at all points. The discontinuity problem comes from the indicator function $\imath(\cdot)$, which allocates a weight of one to all the observations that belong to the interval centered on y and zero weight to the other observations. The principle of kernel estimation is simple: Rather than giving all observations in the interval the same weight, the allocated weight is greater the closer the observation is to y. The

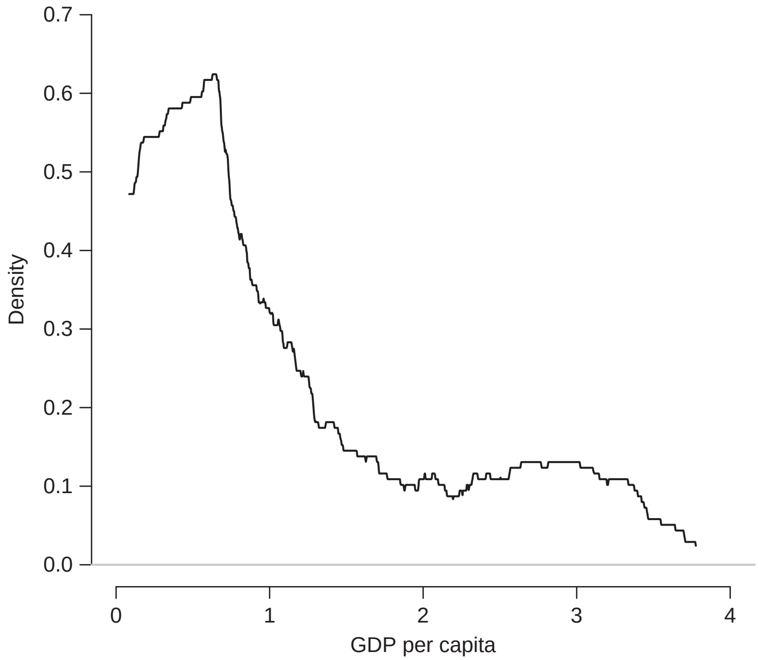


Figure 6.8 Naive estimator of GDP per capita.

transition from 1 to 0 in the weights is then carried out gradually, rather than abruptly. The kernel estimator is obtained by replacing the indicator function by a kernel function $K(\cdot)$:

$$\hat{f}(y) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{y - y_i}{h}\right). \tag{6.13}$$

For $\hat{f}(y)$ to conserve the properties of a density function, the integral of the kernel function over the entire space has to be equal to one. Any probability distribution can then be used as kernel function. The Gaussian and the Epanechnikov distributions are two kernels commonly used in practice.[18] Kernel density estimation is known to be sensitive to the choice of the bandwidth h, whereas it is not really affected by the choice of the kernel function.

[18] The Gaussian kernel corresponds to the choice of the standard Normal distribution: $K(x) = e^{-x^2/2}/\sqrt{2\pi}$. Epanechnikov (1969) proposed a second-degree polynomial, adjusted to satisfy the properties of a density function: $K(x) = 3(1 - x^2/5)/\left(4\sqrt{5}\right)$ if $|x| < \sqrt{5}$ and 0 otherwise.

6.3.2.2 Bandwidth Selection

The question of which value of h is the most appropriate is particularly a thorny one, even if automatic bandwidth selection procedures are often used in practice. Silverman's rule of thumb is mostly used, which is defined as follows:[19]

$$\hat{h}_{\text{opt}} = 0.9 \min\left(\hat{\sigma}; \frac{\hat{q}_3 - \hat{q}_1}{1.349}\right) n^{-1/5}, \tag{6.14}$$

where $\hat{\sigma}$ is the standard deviation of the data, and $\hat{q}_3$ and $\hat{q}_1$ are, respectively, the third and first quartiles calculated from the data. This rule boils down to using the minimum of two estimated measures of dispersion: the variance, which is sensitive to outliers, and the interquartile range. It is derived from the minimization of an approximation of the mean integrated squared error (MISE), a measure of discrepancy between the estimated and the true densities, where the Gaussian distribution is used as a reference distribution. This rule works well in numerous cases. Nonetheless, it tends to oversmooth the distribution when the true density is far from the Gaussian distribution, for example if it is multimodal and highly skewed. Figure 6.2 is a kernel density estimation of GDP per capita with Silverman's rule-of-thumb bandwidth selection. It appears as a smoothed version of the naive estimator in Figure 6.8. Several other data-driven methods for selecting the bandwidth have been developed such as cross-validation (CV) (Bowman, 1984; Rudemo, 1982; Stone, 1974) and plug-in methods (Ruppert et al., 1995; Sheather and Jones, 1991), among others.

Rather than using a Gaussian reference distribution in the approximation of the MISE, the *plug-in* approach consists of using a prior nonparametric estimate, and then choosing the h that minimizes this function. This choice of bandwidth does not then produce an empirical rule as simple as that proposed by Silverman, as it requires numerical calculation. For more details, see Sheather and Jones (1991).

Rather than minimizing the MISE, the underlying idea of CV *by least squares* is to minimize the integrated squared error (ISE). In other words, we use the same criterion, but not expressed in terms of expectations. The advantage of the ISE criterion is that it provides an optimal formula for h for a given sample. The counterpart is that two samples drawn from the same density will lead to two different optimal bandwidth choices. The ISE solution consists in finding the value of h that minimizes: $\text{ISE}(h) = \int \left[\hat{f} - f\right]^2 dy = \int \hat{f}^2 dy - 2\int \hat{f} f\, dy + \int f^2 dy$, where, for simplicity, f and $\hat{f}$ correspond to $f(y)$ and $\hat{f}(y)$. The last term in this equation does not contain h and thus plays no role in the minimization. Furthermore, the term $\int \hat{f} f\, dy$ is exactly $E\left(\hat{f}\right)$. Let $\hat{f}_{-i}$ be the estimator of the density based on the sample containing all the observations except for y_i.

[19] See Equation (3.31) in Silverman (1986, p. 48).

An unbiased estimator of $E(\hat{f})$ is given by $n^{-1}\sum_{i=1}^{n}\hat{f}_{-i}$. The minimization of the ISE criterion thus requires us to minimize the following expression:

$$\mathrm{CV}(h) = \int \hat{f}^2(y)\mathrm{d}y - \frac{2}{n}\sum_{i=1}^{n}\hat{f}_{-i}(y_i).$$

This method is also called *unbiased* CV, as $\mathrm{CV}(h) + \int f^2 \mathrm{d}y$ is an unbiased estimator of MISE. The value of h that minimizes this expression converges asymptotically to the value that minimizes the MISE.

6.3.2.3 Adaptive Kernel Estimator

In the kernel density estimation presented earlier, the bandwidth remains constant at all points where the distribution is estimated. This constraint can be particularly onerous when the concentration of data is markedly heterogeneous in the sample. There would be advantages to using narrower bandwidths in dense parts of the distribution (the middle) and wider ones in the more sparse parts (the tails). The adaptive kernel estimator is defined as follows:

$$\hat{f}(y) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h\lambda_i}K\left(\frac{y-y_i}{h\lambda_i}\right),$$

where λ_i is a parameter that varies with the local concentration of the data. An estimate of the density at point y_i, denoted by $\widetilde{f}(y_i)$, measures the concentration of the data around this point: a higher value of $\widetilde{f}(y_i)$ denotes a greater concentration of data, whereas smaller values indicate lighter concentrations. The parameter λ_i can thus be defined as being inversely proportional to this estimated value: $\lambda_i = \left[g/\widetilde{f}(y_i)\right]^{\theta}$, where g is the geometric mean of $\widetilde{f}(y_i)$ and θ is a parameter that takes on values between 0 and 1.[20] The parameter λ_i is smaller when the density is greater (notably toward the middle of the distribution) and larger when the density is lighter (in the tails of the distribution).

Figure 6.9 presents the adaptive kernel density estimation of GDP per capita across different countries in 1988. Compared to the simple kernel density estimation, with fixed bandwidth (dashed line), the first mode is higher, and the second mode lower.

Several empirical studies on income distribution have used kernel density estimation, among others, see Marron and Schmitz (1992), Jenkins (1995), Cowell et al. (1996), Daly et al. (1997), Quah (1997), Burkhauser et al. (1999), Bourguignon and Morrisson (2002), Pittau and Zelli (2004), Jenkins and Van Kerm (2005), and Sala-i-Martin (2006).

[20] In practice, an initial fixed-bandwidth kernel estimator can be employed as $\widetilde{f}(y_i)$, with $\theta = 1/2$ and λ obtained with Silverman's rule of thumb.

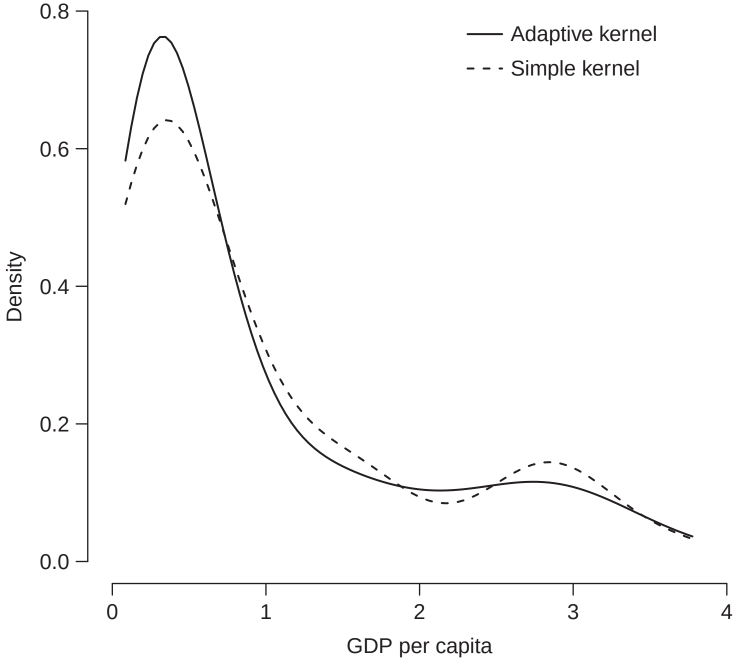


Figure 6.9 Adaptive kernel estimation of GDP per capita.

6.3.2.4 Multivariate and Conditional Density

The extension to the multivariate case is straightforward. The joint density of two variables y and x, for which we have n observations, can be estimated with a bivariate kernel function

$$\hat{f}(y, x) = \frac{1}{nh_1 h_2} \sum_{i=1}^{n} K\left(\frac{y_i - y}{h_1}; \frac{x_i - x}{h_2}\right), \tag{6.15}$$

which is equivalent to the product of two univariate kernels in the Gaussian case. The extension to the d-dimensional case is immediate, via the use of multivariate kernels in d-dimensions. Scott (1992) extends Silverman's rule of thumb as follows: $h_j = n^{-1/(d+4)}\hat{\sigma}_j$, where $\hat{\sigma}_1$ and $\hat{\sigma}_2$ are the sample standard deviations of, respectively, y and x. In practice, kernel density estimation is rarely used with more than two dimensions. With three or more dimensions, not only may the graphical representation be problematic, but the precision of the estimation may be also. Silverman (1986) showed that the number of observations required to guarantee a certain degree of reliability rises explosively with the number of dimensions. This problem is known as the *curse of dimensionality*.

A conditional density function is equal to the ratio of a joint distribution and a marginal distribution, $f(y|x) = f(x, y)/f(x)$. A kernel conditional density estimation is then given by

$$\hat{f}(y|x) = \frac{\frac{1}{h_1 h_2}\sum_{i=1}^{n} K\left(\frac{y_i - y}{h_1}; \frac{x_i - x}{h_2}\right)}{\frac{1}{h_3}\sum_{i=1}^{n} K\left(\frac{x_i - x}{h_3}\right)}. \tag{6.16}$$

When several conditional variables are considered, the bandwidth selection obtained by CV can mitigate the curse of dimensionality problem if some of them are irrelevant (Fan and Yim, 2004; Hall et al., 2004). Several recent studies have focused on conditional analysis in a nonparametric framework. To evaluate policy effects, the impact of a counterfactual change in the distribution of some covariates on the unconditional distribution of some variable of interest has been investigated in DiNardo et al. (1996), Donald et al. (2000), Chernozhukov et al. (2009), Rothe (2010), and Donald et al. (2012). For more details on kernel density estimation, see Silverman (1986), Paul (1999), Li and Racine (2006), and Ahamada and Flachaire (2010).

6.3.3 Finite Mixture Models

6.3.3.1 A Group Decomposition Approach

A population can be decomposed into several distinct groups in many different ways. The density function of the population is then equal to the sum of the densities associated with each of the different groups. Let us consider κ groups, with the density function of each group being parametric, $f_k(y; \theta_k)$ for $k = 1, \ldots, \kappa$ where θ is a set of parameters. Then, the density function of the population can be written,

$$f(y; \theta) = \sum_{k=1}^{\kappa} \pi_k f_k(y; \theta_k) \tag{6.17}$$

where π_k is the proportion of the population belonging to subgroup k. The conditions $0 \le \pi_k \le 1$ and $\sum_{k=1}^{K} \pi_k = 1$ are required to guarantee that the population density integrates to one over the support. A density estimation by mixture models is obtained by replacing the unknown parameters by estimated parameters. In finite mixture models, the group to which each individual belongs is not observed.[21] They thus allow us to capture the effect of unobserved heterogeneity. They can also be used for classification purpose. Bayes' theorem allows us to deduce the *a posteriori* probability that an observation i belongs to the group k:

[21] When the groups are known and also the densities associated to each groups, the mixture model is entirely parametric and can be estimated by maximum likelihood (see Section 6.3.1).

$$\pi_{ik} = \frac{\pi_k f_k(y_i;\theta_k)}{\sum_{k=1}^{\kappa} \pi_k f_k(y_i;\theta_k)}. \tag{6.18}$$

Replacing the unknown parameters by consistent estimates, these individual probabilities can be used to classify the observations into the different groups.

The estimation of a density by a mixture model allows us to bring out the link between parametric and nonparametric estimation. If we consider one single group ($\kappa = 1$), then the mixture models amount to just one parametric function. Adding additional groups allows us to estimate more complicated densities, which cannot be modeled with one sole group; adding more groups allows us to reflect the heterogeneity of the population. Mixture models thus permit much greater modeling flexibility. In the extreme case, where we have as many groups as we do observations ($\kappa = n$), the mixture is equivalent to the estimation of a density by kernel methods (see Section 6.3.2).[22] For values of κ between 1 and the size of the sample n, the mixture model can thus be seen as a semiparametric compromise between parametric estimation and nonparametric kernel estimation. The parametric aspect is reflected in the fact that the density is expressed as a sum of parametric density functions; the nonparametric aspect is captured by the presence of a number of different groups.

The theory of mixture models tells us that, under regularity conditions, any probability density can be consistently estimated as a mixture of Normal distributions.[23] Figure 6.10 depicts a number of different mixtures of two Normal distributions, of which the density can be written as $\pi_1\phi(y;\mu_1,\sigma_1)+\pi_2\phi(y;\mu_2,\sigma_2)$, where $\phi(\cdot)$ is the density of the Normal distribution, with mean μ_k and variance σ_k, for $k = 1, 2$. The global density and the two individual components are represented in the same figure.[24] From global densities (solid lines), we can see that a wide variety of densities can be represented by a mixture of only two Normal distributions, as top flat (panel a), bimodal (panel b), skewed (panel c), and thick upper-tailed (panel d) distributions. Many further examples can be provided to illustrate the very wide variety of distributions that can be characterized by a mixture of κ Normal distributions: see, among others, Marron and Wand (1992). All these examples reveal the great flexibility of finite mixture models in estimating densities.

[22] If $K = n$ and $\pi_1 = \cdots = \pi_K = 1/n$, Equation (6.17) is then equivalent to Equation (6.13) where the $f_k(\cdot)$ function is the kernel function $K(\cdot)$.

[23] See Escobar and West (1995), Ferguson (1983), Titterington, Makov, and Smith (1985), McLachlan and Peel (2000), and Ghosal and van der Vaart (2001).

[24] In panel (a), $\pi_1 = \pi_2 = 0.5$, $\sigma_1 = \sigma_2 = 1$, $\mu_1 = 0$, $\mu_2 = 2$. In panel (b), $\pi_1 = \pi_2 = 0.5$, $\sigma_1 = \sigma_2 = 1$, $\mu_1 = 0$, $\mu_2 = 4$. In panel (c), $\pi_1 = 0.6$, $\pi_2 = 0.4$, $\sigma_1 = \sigma_2 = 1$, $\mu_1 = 0$, $\mu_2 = 2$. In panel (d), $\pi_1 = 0.75$, $\pi_2 = 0.25$, $\sigma_1 = 1$, $\sigma_2 = 2$, $\mu_1 = 0$, $\mu_2 = 2$.

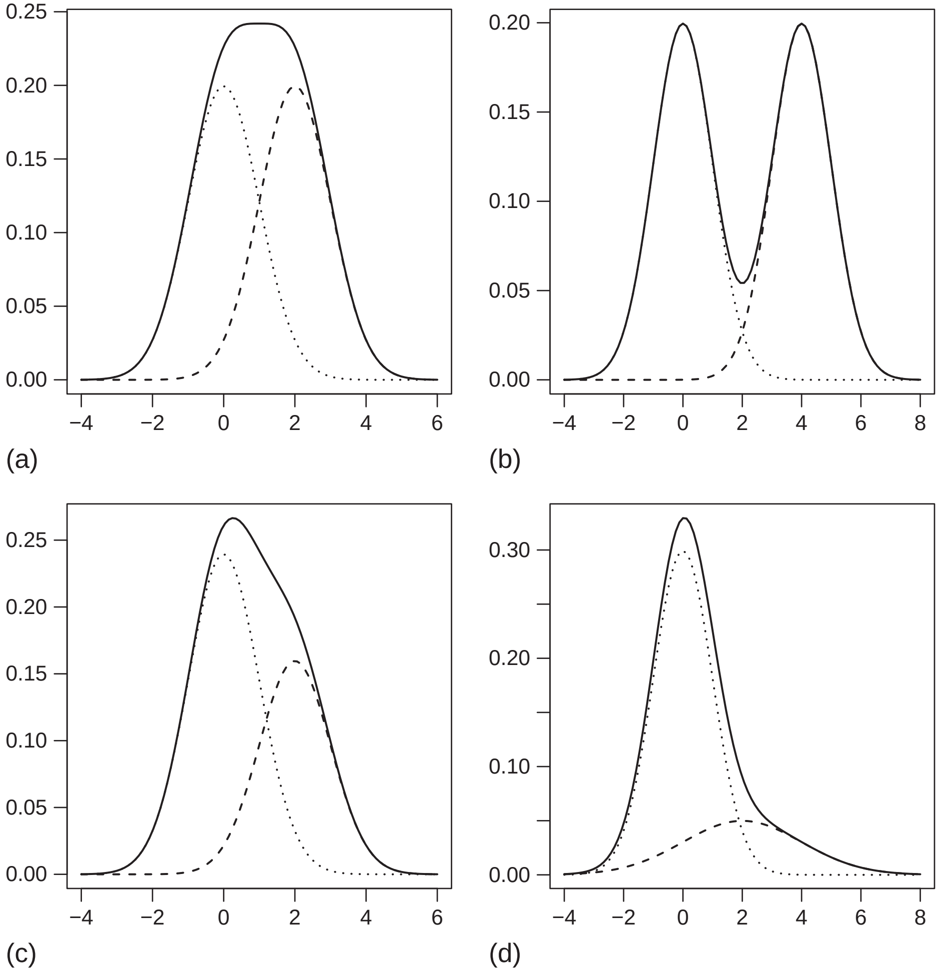


Figure 6.10 Mixtures of two Normal distributions.

6.3.3.2 Number of Components and Number of Groups

For a given number of κ, we can estimate the unknown parameters by maximum likelihood.[25] The number κ, known as the number of components, can be selected by minimizing a criterion, as the Bayesian Information Criterion (BIC),

$$\mathrm{BIC} = -2\hat{\ell} + \#\mathrm{param}\log n, \tag{6.19}$$

[25] The EM algorithm of Dempster et al. (1977) is often used. Bayesian methods can also be employed, see Robert and Casella (2005) and Frühwirth-Schnatter (2006).

where $\hat{\ell}$ is the estimated log-likelihood, #param is the number of parameters to estimate, and n is the number of observations. If the main concern is the best fit of the overall density, this selection criterion is appropriate. However, if the main concern is the detection of distinct groups, the choice of κ is less simple. Indeed, there is no automatic correspondence between the choice of κ and the number of underlying groups in the population. For instance, panel (d) in Figure 6.10 shows that the second component is required to fit a thick upper tail, but it does not clearly identify a distinct group from the first component. Indeed, the two distributions of the groups intersect a lot. Here, the number of component κ is not necessarily equivalent to the number of groups. It illustrates that the definition of what constitutes a distinct group and its detection can be a difficult task in finite mixture models.[26]

Figure 6.11 shows kernel density estimation (on the left) and estimation by a mixture of lognormal distributions (on the right) of the income distribution in the

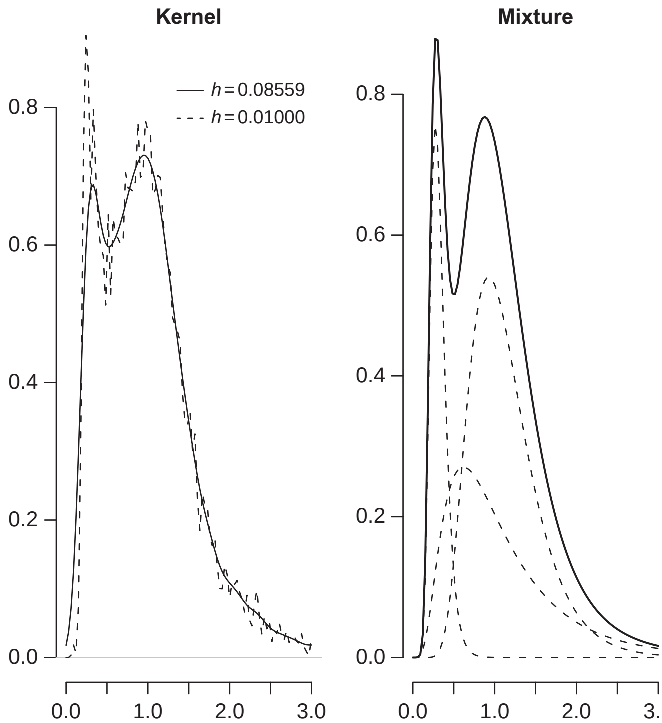


Figure 6.11 Income distribution in the United Kingdom in 1973 (incomes normalized by contemporary mean).

[26] A more appropriate method could be to test the number of modes of the distribution, see Ray and Lindsay (2005).

United Kingdom in 1973. A lognormal distribution is heavy tailed, in the sense that it has an upper tail that is heavier than an exponential. Then, it is more appropriate to use finite mixtures of lognormals rather than finite mixture of normals to fit income distributions, which are typically heavy tailed. The estimation of the density by a mixture of lognormal distributions is obtained from an estimation of the density of log-incomes by a mixture of Normal distributions.[27] The value of the bandwidth given by Silverman's empirical rule ($h=0.08559$) allows us to reproduce the kernel density estimation results in Marron and Schmitz (1992). Kernel estimation with a smaller value of $h(0.01)$ are overlaid in the same figure. The comparison of the two estimators reveals that the results differ significantly: with $h=0.08559$, the first mode is smaller than the second, whereas with $h=0.01$ the reverse holds. This confirms that the kernel estimation with Silverman's rule of thumb does indeed tend to oversmooth the function when the underlying distribution is multimodal and highly skewed distribution (see Section 6.3.2). In our example, the Silverman selection choice tends to flatten out considerably the first mode relative to the second. The right panel shows the density estimation with a mixture of lognormal distributions, obtained by minimizing the BIC. The overall distribution appears to be a smoothed representation of the kernel density estimation with $h=0.01$. In addition, the mixture estimation identifies three separate components. The first and the third components do not overlap a lot; they can be associated to two distinct modes. The second component overlaps to a considerable extent with the third and to a lesser extent with the first. The presence of this second component allows a better fit of the right-hand side tail of the distribution, but cannot be clearly associated to a distinct group.

A very few empirical studies have used finite mixture models to estimate income distributions. Flachaire and Nuñez (2007) studied the distribution of household income in the United Kingdom with a mixture of lognormal distributions. Pittau and Zelli (2006) and Pittau et al. (2010) studied the evolution of per capita income distributions across EU regions and countries. Chotikapanich and Griffiths (2008) estimated the Canadian income distribution using a mixture of Gamma distributions. Lubrano and Ndoye (2011) modeled the income distribution using a Bayesian approach and a mixture of lognormal densities.[28]

6.3.3.3 Group Profiles Explanation

In addition to the estimate of a density function of any form, finite mixture estimation can be used to explain the profiles of the different groups underlying the overall population. This can be done by introducing covariates in the probabilities π_k:

[27] From $f(x;\Theta)=\sum_{k=1}^{K}\pi_k\phi(x;\mu_k,\sigma_k)$, we have $f(y;\Theta)=\sum_{k=1}^{K}\pi_k\Lambda(y;\mu_k,\sigma_k)$ where $x=\log y$.

[28] Paap and Van Dijk (1998) considered mixtures of two distributions, using Normal, lognormal, Gamma, and Weibull distributions. However, their approach is entirely parametric, with the number of components and the densities of each groups fixed *a priori*.

$$f(y|z;\Theta) = \sum_{k=1}^{K} \pi_k(z, \alpha_k) f_k(y; \theta_k), \tag{6.20}$$

where $z = \{z_1, \ldots, z_l\}$ is a vector of l observed variables and $\alpha_k = \{\alpha_{1k}, \ldots, \alpha_{lk}\}$ is a vector of l unknown parameters. This model defines a conditional density function, which takes into account directly the fact that the probability of group membership may be a function of individual characteristics (a white-collar worker has a greater probability of belonging to the group of the richest households than does a blue-collar worker). As well as the nonparametric estimation of the density and the decomposition into different groups, covariates also explain the variability between groups. The relationship between the probabilities π_k and the covariates z can be specified with an ordered logit/probit or multinomial regression model, and the unconditional density can be obtained as follows:

$$f(y; \Theta) = \sum_{k=1}^{K} \overline{\pi}_k \phi(y; \mu_k, \sigma_k) \quad \text{with} \quad \overline{\pi}_k = \frac{1}{n} \sum_{i=1}^{n} \pi_k(z_i, \alpha_k),$$

where z_i represents the vector of characteristics of the ith observation, and n is the number of observations. In other words, $\pi_k(z_i, \alpha_k)$ is the probability that the individual i with characteristics z_i belong to the group k. For more details, see Ahamada and Flachaire (2010).

Figure 6.12 reproduces the results of the estimation of the distribution of household income in the United Kingdom in 1979 and 1988, obtained in Flachaire and Nuñez (2007), by a mixture of lognormal distributions.[29] The decomposition into groups of the mixture estimator does emphasize clear changes over time that would be difficult to see from the comparisons of the overall distributions. The analysis by groups shows that, in 1988, a small separate group had formed to the extreme left of the distribution, whereas that situated to the far right of the distribution had grown in size. Table 6.3 reproduces the estimated coefficients (with standard errors in parentheses) associated to the following covariates: z_1 for a retired household, z_2 for single-parent families,

Table 6.3 Coefficient estimates of covariates

	z_1	z_2	z_3	z_4	z_5
1979	−1.77 (0.059)	−0.67 (0.106)	0.61 (0.050)	−1.16 (0.086)	−0.44 (0.020)
1988	−1.33 (0.058)	−0.69 (0.106)	0.78 (0.053)	−1.44 (0.068)	−0.35 (0.022)

[29] The analysis here uses the same data as Marron and Schmitz (1992), with the exception that the incomes are normalized via an equivalence scale in order to account for differences in household size.

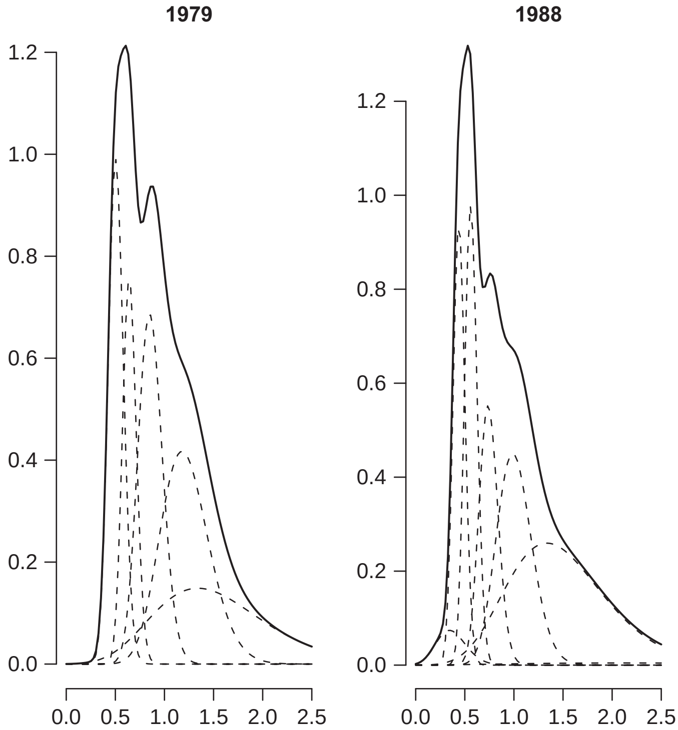


Figure 6.12 Income distribution in the United Kingdom in 1979 and 1988 (incomes normalized by contemporary mean).

z_3 for households where all the adults work, z_4 if there is no adult working in the household (in a nonretired household), and z_5 for the number of children.

An ordered probit model is used to specify the relationship between the probabilities and the covariates. If a coefficient is positive (negative), then the position of an observation with the associated variable moves to the right (left) of the distribution as the variable z_l increases. On the other hand, a value of $\alpha_{l,}$ that is not significantly different from zero, indicates that the characteristic z_l does not help us to explain the decomposition of the sample into the different groups. From the results, we can see that retired (z_1) and nonworking (z_4) households are more likely to be found toward the bottom of the income distribution, and households where all adults work (z_3), on the contrary, are more likely to be found toward the right of the distribution. In addition, the position of retired households has improved over this period, whereas that of households where no one works has deteriorated. These results emphasize the usefulness of mixture models, which yield an overall picture of the distribution of income and how this has changed over time, with richer results than those obtained from other commonly employed techniques.

6.3.3.4 Finite Mixture of Regressions

Covariates have been introduced in the probabilities to characterized group profiles. They can be also introduced into the modeling of the densities in each of the groups, leading us to consider mixtures of regression models. Let us consider a mixture of Normal distributions with variance σ_k^2 and mean being conditional on some covariates, $\mu_k = x\beta_k$, which can be written as:

$$f(y|x;\Theta) = \sum_{k=1}^{\kappa} \pi_k \phi(y|x;\beta_k, \sigma_k). \tag{6.21}$$

If there are two groups ($\kappa = 2$), one just needs to consider the following model:

$$\begin{aligned} &\text{Group 1}: y = x\beta_1 + \varepsilon_1, \quad \varepsilon_1 \sim N\left(0, \sigma_1^2\right), \\ &\text{Group 2}: y = x\beta_2 + \varepsilon_2, \quad \varepsilon_2 \sim N\left(0, \sigma_2^2\right), \end{aligned}$$

where ϵ_1 and ϵ_2 are independent and identical Normally distributed error terms within each group, with variances of σ_1^2 and σ_2^2, respectively. In this model, we consider that the population is composed of two different groups, for which the relationship between the dependent and explanatory variables is different, and the observations come from the different groups in the population in unknown proportions. This specification would be particularly appropriate if we assume that the marginal impact of covariates may be different in each of the groups, as suggested in Figure 6.3. Covariates could be introduced at the same time in the probabilities to explain group profiles.

To illustrate, consider a simple Mincer earnings equation, which explains the logarithm of an individual's earnings by their number of years of education and number of years of labor-market experience. One way of testing for earnings differences between men and women is to test if the parameters of the earnings equation are statistically significantly different between the two groups of individuals, via a Chow test (Chow, 1960). Table 6.4 shows ordinary least squares (OLS) estimation results from linear regression models for the groups of men and of women (columns 1 and 2), using data from a household survey carried out by the U.S. Census Bureau in May 1985.[30] A Chow test, equal to 14.19, rejects the null hypothesis that the two sets of coefficients are identical. As the dependent variable is the log of earnings, the estimation results show that one additional year of education increases earnings by around 7.9% for men and 10.9% for women on average. The earnings profiles as a function of labor-market experience are different between the two groups. These are traced out in Figure 6.13 for 8 years of education, from which we can see that the gender gap is sharply increasing with labor-market experience during the first 30 years.[31]

[30] The data come from Chapter 5 of Berndt (1990).

[31] The curve for the group of men corresponds to the polynomial $y = 0.66194 + 8 \times 0.07941 + 0.04484 \times x - 0.00066 \times x^2$ and that for the group of women to $y = 0.22254 + 8 \times 0.10915 + 0.02597 \times x - 0.00038 \times x^2$. The gap between the two curves widens with experience at first before narrowing again after around 30 years of labor-market experience.

Table 6.4 Mincer earnings equations

	Linear models		Mixture model	
Variables	**Men**	**Women**	**Group 1 ($\pi_1 = 0.406$)**	**Group 2 ($\pi_2 = 0.594$)**
Explanatory variables				
Constant	0.66194⋆	0.2225	0.67517⋆	0.34909⋆
Education	0.07941⋆	0.10915⋆	0.08202⋆	0.09844⋆
Experience	0.04484⋆	0.02597⋆	0.05147⋆	0.02590⋆
(Experience)^2	−0.00066⋆	−0.00038⋆	−0.00078⋆	−0.00040⋆
Concomitant variables				
Constant			–	−0.49423⋆
Female			–	9.03215⋆
Union member				−8.16128†

⋆$p < 0.05$, $^\dagger p = 0.057$

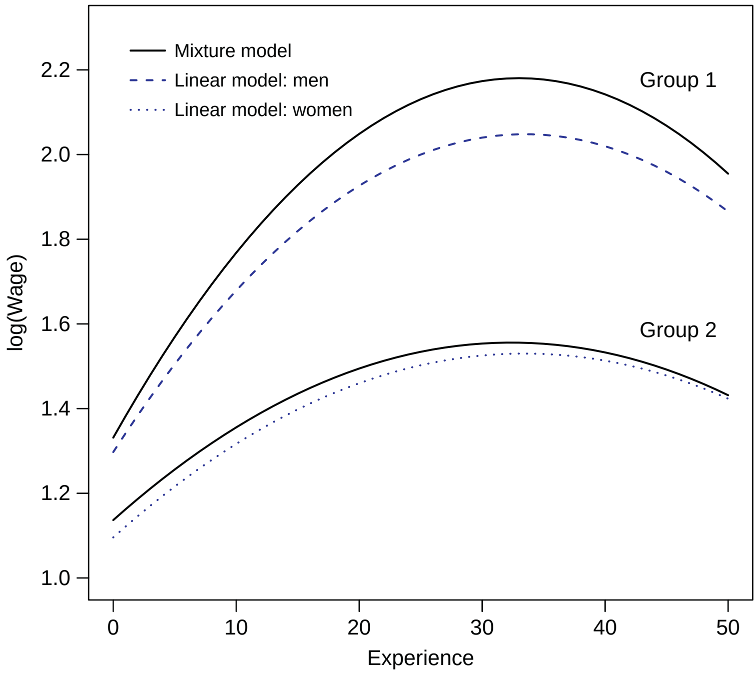


Figure 6.13 The relationship between earnings and labor-market experience.

In the linear model approach, we define *a priori* two groups of individuals—men and women. By contrast, in a mixture model approach, we do not specify the groups *a priori*, we let the data identify homogeneous groups with respect to the relationship between the dependent and explanatory variables. Table 6.4 shows estimation results from a mixture model (columns 3 and 4); the BIC suggests that there are two groups. The estimation results show that one additional year of education increases earnings by around 8.2% for individuals in the first group and 9.8% for individuals in the second, on average. The impact of experience on earnings are traced in Figure 6.13 for 8 years of education (solid lines).[32] From this figure, the gap between the two groups is much larger than the gap between the groups of men and of women obtained from linear models.

The use of concomitant variables in the mixture model allows us to characterize the profile of the groups. Two dummy variables are taken into account as concomitant variables. The first is for the individual being a woman (Female), and the second for the individual being a union member (Union Member). In Table 6.4, column 4, the positive and significant coefficient on the "Female" variable indicates that women are more likely to belong to the second group than to the first group. The negative significant coefficient on "Union Member" shows that unionized workers are less likely to belong to the second than to the first group.[33] A classification shows that 96.3% of women belong to group 2, whereas the analogous percentage of men is only 19%.[34] Equally, the percentage of union members classified in group 1 is 80.2%. Last, the results of this analysis suggest that, for the vast majority of women, the relationship between earnings and experience is much flatter than that for most men and union members, holding everything else equal. The gap obtained is much larger than those obtained by considering all men and all women in two distinct groups.

For more details on mixture models of regression, see McLachlan and Peel (2000), Frühwirth-Schnatter (2006), and Ahamada and Flachaire (2010).

6.3.4 Finite Sample Properties

In this section, we study the quality of the fit of nonparametric density estimation in finite samples. To asses the quality of the density estimation, we need to use a distance measure between the estimated density and the true density. We use the mean integrated absolute errors (MIAE) measure,

[32] The curve for the first group corresponds to the polynomial $y = 0.67517 + 8 \times 0.08202 + 0.05147 \times x - 0.00078 \times x^2$, and that of the second to $y = 0.34909 + 8 \times 0.09844 + 0.02590 \times x - 0.0004 \times x^2$.

[33] This coefficient is significant at the 1% level according to an LR test.

[34] An individual is assigned to a group when his individual *a posteriori* probability of belonging to this group is higher than the probabilities to belong to other groups, see (6.18).

$$\mathrm{MIAE} = E\left(\int_0^{\infty} \left|\hat{f}(y) - f(y)\right| \mathrm{d}y\right). \tag{6.22}$$

In our experiments, data are generated from two unimodal distributions: lognormal, $\Lambda(y;\ 0,\ \sigma)$, and Singh–Maddala distributions, $\mathrm{SM}(y;\ 2.8,\ 0.193,\ q)$. We also use a bimodal distribution: a mixture of two Singh–Maddala distributions: $\frac{2}{5}\mathrm{SM}(y; 2.8, 0.193, 1.7) + \frac{3}{5}\ \mathrm{SM}(y; 5.8, 0.593, q)$, plotted in Figure 6.14. As σ increases and q decreases, the upper tail of the distribution decays more slowly. The sample size is $n = 500$, and the MIAE criterion is calculated as the average of $\int_0^{\infty} |f(y) - f(y)| \mathrm{d}y$ computed for 1000 samples.

Table 6.5 shows the quality of the fit for several density estimation methods. We first consider standard kernel estimator, with fixed bandwidth selected by the Silverman rule-of-thumb (Silv.), by cross-validation (CV) and by the plug-in (Plug-in) methods. Then, we consider adaptive kernel methods based on each of the previous fixed bandwidths. Finally, we consider density estimation based on mixture of lognormal distributions.[35]

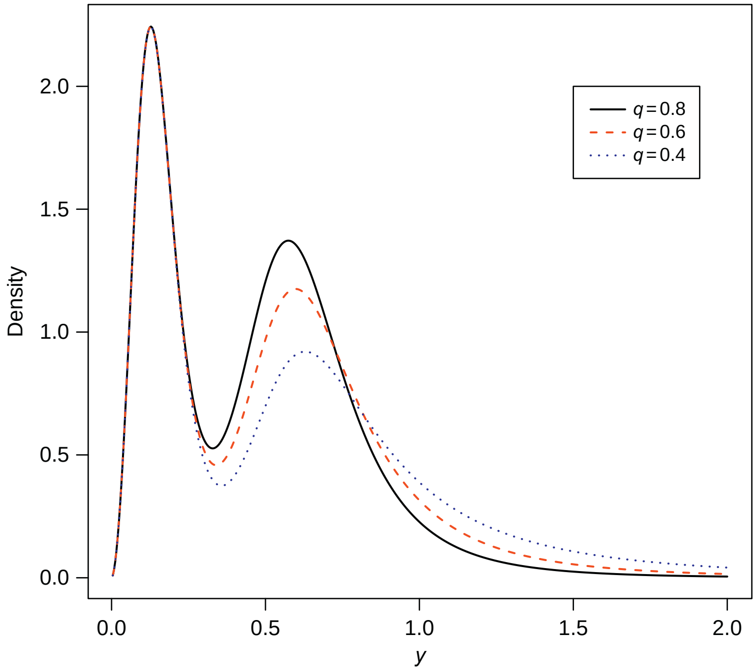


Figure 6.14 Mixture of two Singh–Maddala distributions.

[35] The density function of the logarithmic transformation of the data is estimated by a mixture of normal distributions and the number of components is selected with the BIC.

Table 6.5 Quality of density estimation (MIAE), $n = 500$

	Standard kernel			Adaptive kernel			Mixture lognormal
	Silv.	**CV**	**Plug-in**	**Silv.**	**CV**	**Plug-in**	
Lognormal							
$\sigma = 0.5$	0.1044	0.1094	0.1033	0.0982	0.1098	0.1028	0.0407
$\sigma = 0.75$	0.1326	0.1326	0.1252	0.1098	0.1283	0.1179	0.0407
$\sigma = 1$	0.1643	0.1716	0.1522	0.1262	0.1609	0.1362	0.0407
Singh–Maddala							
$q = 1.7$	0.0942	0.1009	0.0951	0.0915	0.0994	0.0934	0.0840
$q = 1.2$	0.1039	0.1100	0.1048	0.0947	0.1050	0.0994	0.0920
$q = 0.7$	0.1346	0.1482	0.1326	0.1049	0.1349	0.1175	0.0873
Mixture of two Singh–Maddala							
$q = 0.8$	0.2080	0.1390	0.1328	0.1577	0.1356	0.1224	0.1367
$q = 0.6$	0.2458	0.1528	0.1463	0.1896	0.1457	0.1293	0.1464
$q = 0.4$	0.2885	0.1953	0.1733	0.2234	0.1812	0.1450	0.1366

The results in Table 6.5 show that, for standard and adaptive kernel methods, the quality of the fit deteriorates as the upper tail becomes heavier (as σ increases and q decreases, MIAE increases). Moreover, standard kernel method with the Silverman's rule-of-thumb bandwidth fails when the distribution is multimodal and highly skewed (case of two Singh–Maddala distributions), compared to other methods. Finally, our results suggest that, in the cases of heavier-tailed distributions, the adaptive kernel based on the plug-in bandwidth and the mixture of lognormals perform better than standard kernel methods.

6.4. WELFARE INDICES

We can use the term *welfare indices* to cover a number of specific tools of distributional analysis that are of interest to economists and social scientists. These include social welfare functions, inequality measures, and poverty indices. Our approach is to characterize some basic classes of indices, to introduce some standard results that enable us to describe the statistical properties of these indices, and then to apply the analysis to particular welfare indices that are of interest to students of income distribution. The applications here will be to inequality and poverty indices.

6.4.1 Basic Cases

It is useful begin with two of the simplest welfare indices, the quantile and the income cumulation. Quantiles and income cumulations are themselves incomes and so belong to the interval $\mathbb{Y} = [\underline{y}, \overline{y})$, introduced in Section 6.1.2. Once again we work with distribution functions $F \in \mathbb{F}$ so that the total population in the distribution is implicitly normalized to 1.

Let $q \in \mathbb{Q}$ denote an arbitrary population proportion; then we may define Q the *quantile functional* from $\mathbb{F} \times \mathbb{Q}$ to $\mathbb{Y}$ as

$$Q(F; q) := \inf\{y | F(y) \geq q\} \tag{6.23}$$

(Gastwirth, 1971). For any distribution F the quantile functional gives the smallest income in $\mathbb{Y}$ such that $100q$ percent of the population have exactly that income or less. In cases where the distribution F is understood, we can use a shorthand form for the qth quantile

$$y_q := Q(F; q). \tag{6.24}$$

The functional Q provides the basis for several intuitive approaches to the analysis of income distribution. For example, commonly used quantile ratios — such as the "90/10" ratio, the "90/50" ratio (Alvaredo and Saez, 2009; Autor et al., 2008; Burkhauser et al., 2009)—are found by taking pairs of instances of (6.23) with appropriate q-values: y_{90}/y_{10}, y_{90}/y_{50} and so on.

Likewise C, the *cumulative income functional* from $\mathbb{F} \times \mathbb{Q}$ to $\mathbb{Y}$ is defined as

$$C(F; q) := \int_{\underline{y}}^{y_q} y \mathrm{d}F(y) \tag{6.25}$$

(Cowell and Victoria-Feser, 2002); for any distribution F the cumulative functional gives the total income received by the bottom $100q$ percent of the population. Again, in cases where the distribution F is understood, we can use the shorthand form for the qth cumulation: $c_q := C(F; q)$. A word of caution here: remember that the population is normalized to one; this convention is also embedded in the income cumulations (6.25). In particular, if we set $q = 1$ in (6.25), we get

$$c_1 = C(F; 1) = \mu(F), \tag{6.26}$$

the mean of the distribution F. We can find other intuitive approaches to the analysis of income distribution using C: For example, the income share of the poorest $100q$ percent of the population is obtained from two cumulants defined in (6.25) as

$$\frac{c_q}{c_1} = \frac{C(F; q)}{C(F; 1)}. \tag{6.27}$$

However, this is just a beginning. The indices generated by Q and C in (6.23) and (6.25) are but two well-known examples of a large class of welfare indices that can be expressed in *additively decomposable* form

$$W_{\mathrm{AD}}(F) := \int \phi(y) \mathrm{d}F(y), \tag{6.28}$$

up to a transformation involving $\mu(F)$, where $\phi : \mathbb{Y} \times \mathbb{Y} \to \mathbb{R}$ is piecewise differentiable. Decomposability here means decomposable by population subgroups (Cowell and Fiorio, 2011). This property can be seen more intuitively in the special case of a discrete distribution. If F consists of m point masses consisting of mass f_i located at income y_i $i = 1, \ldots, m$ then (6.28) becomes

$$\sum_{i=1}^{m} f_i \phi(y_i). \tag{6.29}$$

It is clear that the form of (6.29) implies that the welfare index can be found by evaluating income y_i in each of the m separate groups, weighting by the population of the group and aggregating.

Of course, it is not only the rather restrictive class W_{AD} that is interesting for distributional analysis. Many welfare indices can be conveniently expressed in the more general *quasi-additively decomposable* form

$$W_{\mathrm{QAD}}(F) := \int \varphi(y, \mu(F)) \mathrm{d}F(y) \tag{6.30}$$

where $\varphi : \mathbb{Y} \times \mathbb{Y} \to \mathbb{R}$ is piecewise differentiable, and most of the other commonly used welfare indices that cannot be expressed in the form (6.30) can be expressed in the *rank-dependent* form

$$W_{\mathrm{RD}}(F) := \int \psi(y, \mu(F), F(y)) \mathrm{d}F(y), \tag{6.31}$$

where ψ is piecewise differentiable. We will discuss specific examples from the W_{QAD} and W_{RD} classes of functionals in Sections 6.4.3 and 6.4.4 on inequality and poverty measures.

6.4.2 Asymptotic Inference

In this section and in Sections 6.4.3 and 6.4.4, we focus on estimation and inference problems for cases where the sample size n may be considered to be arbitrarily large.[36] The small sample problem is discussed in Section 6.4.5. Furthermore, for the moment we will concentrate only on *distribution-free* approaches that do not require any estimation of the density function, parametric, or even nonparametric; the parametric approach is considered in Section 6.4.6.

There are several methods that we can use to derive the tools that we need. Here, we will make extensive use of an approach that enables us to derive the asymptotic results quickly and simply and that lays the basis for further discussions in Section 6.6.[37]

[36] For an overview of literature, see Cowell (1999).
[37] This approach draws heavily on Cowell and Victoria-Feser (2003).

6.4.2.1 The Influence Function

The principal analytical tool employed here is the *influence function* (IF), which can be used here as a device to quantify the effect of a perturbation on some given theoretical distribution. So, assume that $F \in \mathbb{F}$ is the distribution in question and that $H^{(z)} \in \mathbb{F}$ is another distribution that consists just of a single point mass at z,

$$H^{(z)}(y) = \iota(y \geq z), \tag{6.32}$$

where $\iota(\cdot)$ is the indicator function (6.1). Then the mixture distribution

$$G := [1-\delta]F + \delta H^{(z)}, 0 \leq \delta \leq 1 \tag{6.33}$$

can be taken as a representation of the perturbation of the distribution F by the point mass, where δ represents the relative size of the perturbation. Now we need a way of quantifying the importance of this perturbation of F: consider a functional $T : \mathbb{F} \to \mathbb{R}^m$ that represents some statistic in which we are interested. The IF measures the impact of the perturbation on the statistic T for infinitesimal δ, namely,

$$\mathrm{IF}(z; T, F) := \lim_{\delta \downarrow 0} \left[\frac{T(G) - T(F)}{\delta} \right] \tag{6.34}$$

which becomes $\frac{\partial}{\partial \delta} T(G)|_{\delta \to 0}$ if T is differentiable.

The IF is particularly useful in analyzing the problem of data contamination (see Section 6.6.1). But the IF has other convenient applications: its relevance to this part of our discussion is that it may be used to derive asymptotic results such as asymptotic covariance matrices. If the distribution G is "near" F (as in Equation (6.33) for small δ) then the first-order von Mises expansion of T at F evaluated in G is given by

$$T(G) = T(F) + \int \mathrm{IF}(y; T, F)\mathrm{d}(G - F)(y) + \text{remainder}$$

When the observations are independently and identically distributed according to F then, by the Glivenko–Cantelli theorem, the empirical distribution $F^{(n)} \to F$. So we may replace G by $F^{(n)}$ for sufficiently large n and obtain

$$T\left(F^{(n)}\right) \approx T(F) + \frac{1}{n}\sum_{i=1}^{n} \mathrm{IF}(y_i; T, F) + \text{remainder}$$

from which we obtain (Hampel et al., 1986, p. 85):

Lemma 1

When the remainder becomes negligible as $n \to \infty$, by the CLT, $\sqrt{n}\left(T\left(F^{(n)}\right) - T(F)\right)$ is asymptotically normal with asymptotic covariance matrix

$$\int \mathrm{IF}(y; T, F)\mathrm{IF}^{\top}(y; T, F)\mathrm{d}F(y) \tag{6.35}$$

Regularity conditions can be found in Reeds (1976), Boos and Serfling (1980), and Fernholz (1983).

Lemma 1 constitutes the basis of the results that follow. Given a statistic T, one just needs to compute its IF to obtain the asymptotic covariance matrix. For inequality and poverty measures (unidimensional statistics), T is a functional $\mathbb{F} \to \mathbb{R}$. In many cases, we can express the IF as a random variable Z minus its expectation,

$$\mathrm{IF}(y, T, F) = Z - E(Z) \tag{6.36}$$

For unidimensional statistics, from Lemma 1, $\sqrt{n}\left(T\left(F^{(n)}\right) - T(F)\right)$ is then asymptotically normal with asymptotic variance equal to

$$\int \mathrm{IF}(y, T, F)^2 \mathrm{d}F(y) = \int (Z - E(Z))^2 \mathrm{d}F(Z), \tag{6.37}$$

which is nothing but the variance of Z. This result allows us to estimate the asymptotic variance of the statistic from a sample using

$$\widehat{\mathrm{var}}\left(T\left(F^{(n)}\right)\right) = \frac{1}{n}\widehat{\mathrm{var}}(Z) = \frac{1}{n^2}\sum_{i=1}^{n}(Z_i - \overline{Z})^2 \tag{6.38}$$

where Z_i, for $i = 1, \ldots, n$ are sample realizations of Z and $\overline{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$. From a sample, the asymptotic variance of the statistic is simple to estimate; it is the empirical variance of $Z_1, \ldots, Z_n$ divided by n.

The main issue here is to provide IFs and to express them as a function of Z as in (6.36), for a wide range of welfare indices and ranking tools, as well as for different forms of data. Moreover, for some important cases, we will also develop analytically the formula in (6.35) so that the approach for computing asymptotic covariance matrices based on the IF can be compared to those from other approaches in the literature.[38]

6.4.2.2 Background Results

Several useful results in income distribution analysis can be found from a simple application of the IF; in particular, we have two key properties for the fundamental functionals introduced in Section 6.4.1 (Cowell and Victoria-Feser, 2002). Applying (6.23) to the distribution in (6.33), we get the qth quantile in the mixture distribution:

$$Q(G, q) = Q\left(F, \frac{q - \iota\left(y_q \geq z\right)\delta}{1 - \delta}\right) \tag{6.39}$$

[38] For previous suggestions on the use of the IF for estimating asymptotic variances, see, for example, Efron (1982) and Deville (1999).

where $y_q = Q(F, q)$ is the qth quantile for the (unmixed) income distribution. Let f be the density function for the distribution function F; then, differentiating (6.39) with respect to δ and setting $\delta = 0$, we obtain the following result.

Lemma 2

The IF for the quantile functional is

$$\mathrm{IF}(z; Q(\cdot, q), F) = \frac{q - \iota(Q(F; q) \geq z)}{f(Q(F; q))} = \frac{q - \iota(y_q \geq z)}{f(y_q)}. \tag{6.40}$$

Likewise, if we apply (6.25) to the distribution in (6.33), we get the qth income cumulation in the mixture distribution:

$$C(G; q) = [1 - \delta] \int_{\underline{y}}^{Q(G; q)} y \mathrm{d}F(y) + \delta z \tag{6.41}$$

where $Q(G, q)$ is given by (6.39). Once again, differentiating (6.41) with respect to δ and setting $\delta = 0$ we obtain another basic result.

Lemma 3

The IF for the cumulative income functional is

$$\begin{aligned} \mathrm{IF}(z; C(\cdot; q), F) &= qQ(F; q) - C(F; q) + \iota(q \geq F(z))[z - Q(F; q)] \\ &= qy_q - c_q + \iota(y_q \geq z)[z - y_q] \end{aligned}. \tag{6.42}$$

We will find that these results are useful not only for welfare indices considered in this section but also for distributional comparisons treated in Section 6.5.

6.4.2.3 QAD Welfare Indices

Let us first deal with the broad W_{QAD} class, the welfare indices that are quasi-additively decomposable; we will turn to the important, but more difficult, rank-dependent class W_{RD} later. Fortunately, this class covers a great number of commonly used tools of distributional analysis; fortunately also the properties are straightforward. Given a sample $y_1, \ldots, y_n$, the sample analogues of W_{QAD} defined in (6.30) are given by

$$\hat{W}_{\mathrm{QAD}} := W_{\mathrm{QAD}}\left(F^{(n)}\right) = \frac{1}{n} \sum_{i=1}^{n} \varphi(y_i, \hat{\mu}) \tag{6.43}$$

where $F^{(n)}$ is the EDF defined in Equation (6.8) and $\hat{\mu}$ is the sample mean:

$$\hat{\mu} := \mu\left(F^{(n)}\right) = \frac{1}{n} \sum_{i=1}^{n} y_i. \tag{6.44}$$

Substituting the mixture distribution (6.33) into (6.30), differentiating with respect to δ and evaluating at $\delta = 0$, we find the IF for the QAD class as:

$$\mathrm{IF}\left(z;\, W_{\mathrm{QAD}},\, F\right) = \varphi(z,\mu(F)) - W_{\mathrm{QAD}}(F) + [z-\mu(F)]\int \varphi(z,\mu(F))\mathrm{d}F(z) \tag{6.45}$$

where φ_μ denotes the partial derivative with respect to the second argument. This IF can be expressed as in (6.36), that is, as a random variable Z minus its expectation,

$$\mathrm{IF}\left(y,\, W_{\mathrm{QAD}},\, F\right) = Z - E(Z), \tag{6.46}$$

where

$$Z = \varphi(y,\mu(F)) + y\int \varphi_\mu(y,\mu(F))\mathrm{d}F(y). \tag{6.47}$$

From (6.36) and (6.37), the asymptotic variance of $\sqrt{n}\left(\hat{W}_{\mathrm{QAD}} - W_{\mathrm{QAD}}\right)$ is then equal to the variance of Z. From (6.38), the asymptotic variance of $\hat{W}_{\mathrm{QAD}}$ can be estimated from a sample using

$$\widehat{\mathrm{var}}\left(\hat{W}_{\mathrm{QAD}}\right) = \frac{1}{n}\widehat{\mathrm{var}}(Z). \tag{6.48}$$

6.4.3 Application: Inequality Measures

Almost all commonly used inequality indices other than the Gini coefficient can be written in the form $\Psi(W_{\mathrm{QAD}}(F),\mu(F))$ where $\Psi:\mathbb{R}^2 \to \mathbb{R}$. So we can use the results on the broad W_{QAD} class to derive the sampling distribution for a large range of inequality measures. We consider two leading examples in Sections 6.4.3.1 and 6.4.3.2.

6.4.3.1 The Generalized Entropy Class

We consider first an important family of inequality measures that belongs to the additively decomposable class (6.28). Members of the GE class (characterized by the parameter ξ) are defined by Equations (6.49)–(6.51)

$$I_{\mathrm{GE}}^{\xi}(F) = \frac{1}{\xi^2-\xi}\left[\int_{\underline{y}}^{\bar{y}}\left[\frac{y}{\mu(F)}\right]^{\xi}\mathrm{d}F(y) - 1\right], \xi\in\mathbb{R},\quad \xi\neq 0,1 \tag{6.49}$$

$$I_{\mathrm{GE}}^{0}(F) = -\int_{\underline{y}}^{\bar{y}}\log\left(\frac{y}{\mu(F)}\right)\mathrm{d}F(y) \tag{6.50}$$

$$I_{\mathrm{GE}}^{1}(F) = \int_{\underline{y}}^{\bar{y}}\frac{y}{\mu(F)}\log\left(\frac{y}{\mu(F)}\right)\mathrm{d}F(y) \tag{6.51}$$

Clearly the GE class belongs to the decomposable class of indices given in (6.28), a subset of the broad W_{QAD} class. The parameter ξ of the GE class characterizes the sensitivity to income differences in different parts of the income distribution. The more positive (negative) ξ is, the more sensitive is the inequality measure to income differences at the top (bottom) of the distribution. The mean logarithmic deviation (MLD) index, $I^0_{\mathrm{GE}}(F)$, is the limiting case when $\xi = 0$. The Theil index, $I^1_{\mathrm{GE}}(F)$, is the limiting case of the GE when $\xi = 1$. The sample analogues of these indices are given by

$$\hat{I}^{\xi}_{\mathrm{GE}} := I^{\xi}_{\mathrm{GE}}\left(F^{(n)}\right) = \begin{cases} \left[n\left(\xi^2 - \xi\right)\right]^{-1} \sum_{i=1}^{n} \left[(y_i/\hat{\mu})^{\xi} - 1\right] & \text{for } \xi \neq 0, 1 \\ -n^{-1} \sum_{i=1}^{n} \log(y_i/\hat{\mu}) & \text{for } \xi = 0 \\ n^{-1} \sum_{i=1}^{n} (y_i/\hat{\mu}) \log(y_i/\hat{\mu}) & \text{for } \xi = 1 \end{cases} \tag{6.52}$$

Using (6.46)–(6.48), we find that the variance of the GE measures can be estimated by

$$\widehat{\mathrm{var}}\left(\hat{I}^{\xi}_{\mathrm{GE}}\right) = \frac{1}{n^2} \sum_{i=1}^{n} (Z_i - \overline{Z})^2 \tag{6.53}$$

where

$$Z_i = \begin{cases} \left(\xi^2 - \xi\right)^{-1} (y_i/\hat{\mu})^{\xi} - \xi (y_i/\hat{\mu}) \left[\hat{I}^{\xi}_{\mathrm{GE}} + \left(\xi^2 - \xi\right)^{-1}\right] & \text{for } \xi \neq 0, 1 \\ (y_i/\hat{\mu}) \log y_i & \text{for } \xi = 0 \\ (y_i/\hat{\mu}) \left[\log(y_i/\hat{\mu}) - \hat{I}^{1}_{\mathrm{GE}} - 1\right] & \text{for } \xi = 1 \end{cases} \tag{6.54}$$

From a sample $y_1, \ldots, y_n$, the values of $Z_1, \ldots, Z_n$ can be calculated for a fixed ξ. The variance estimate of the generalized index is then computed as the empirical variance of $Z_1, \ldots, Z_n$ divided by n.

To show the results in (6.54), let us consider the case $\xi \neq 0, 1$, from which we have

$$\varphi(y, \mu(F)) = \frac{1}{\xi^2 - \xi} \left[\left[\frac{y}{\mu(F)}\right]^{\xi} - 1\right] \tag{6.55}$$

$$\varphi_{\mu}(y, \mu(F)) = \frac{-\xi}{\xi^2 - \xi} \left[\frac{y^{\xi}}{\mu(F)^{\xi+1}}\right] = -\frac{\xi}{\mu} \left(\varphi(y, \mu(F)) + \frac{1}{\xi^2 - \xi}\right) \tag{6.56}$$

Substituting (6.55) and (6.56) into (6.47) gives the result in (6.54), where y is replaced by its sample realization y_i. The same methodology can be applied for the cases $\xi = 0$ and $\xi = 1$.

Clearly the same approach can be applied to functions of moments of the distribution such as the coefficient of variation. Likewise, the statistical properties of the Atkinson class of inequality indices (Atkinson, 1970)

$$I_{\text{Atk}}^{\xi}(F) = 1 - \left[\int_{\underline{y}}^{\overline{y}} \left[\frac{y}{\mu(F)} \right]^{\xi} \mathrm{d}F(y) \right]^{1/\xi}, \quad \xi < 1 \tag{6.57}$$

can easily be derived from (6.53).[39]

The standard approach to obtain the results for the GE class and associated indices is by expressing the indices as a function of moments of the distribution and using the delta method. We can show that both IF and delta approaches give the same results. Indeed, from (6.28) and (6.29), a decomposable inequality measure can written as a function of two moments,

$$I = \psi(\nu; \mu) \quad \text{with} \quad \mu = E(y) \quad \text{and} \quad \nu = E(\phi(y)), \tag{6.58}$$

where ϕ and ψ are functions $\mathbb{R}^2 \to \mathbb{R}$ and ψ is monotonic increasing in its first argument; in particular, this is true for the I_{GE}^{ξ} and I_{Atk}^{ξ} families. The estimation of inequality indices is usually obtained by replacing the unknown moments of the distribution by consistent estimates. The moments are directly estimated by their sample counterparts. Let us consider y_i, for $i = 1, \ldots, n$, a sample of IID observations drawn from F. The estimator of the inequality measure can be expressed as a nonlinear function of two consistently estimated moments,

$$\hat{I} = \psi(\hat{\nu}; \hat{\mu}) \quad \text{with} \quad \hat{\mu} = \frac{1}{n}\sum_{i=1}^{N} y_i \quad \text{and} \quad \hat{\nu} = \frac{1}{n}\sum_{i=1}^{N} \phi(y_i). \tag{6.59}$$

From the CLT, this estimator is also consistent and asymptotically Normal, with asymptotic variance that can be calculated by the delta method. Specifically, the asymptotic variance is equal to

$$\operatorname{var}(\hat{I}) = \left(\frac{\partial \psi}{\partial \nu} \right)^2 \operatorname{var}(\hat{\nu}) + 2 \left(\frac{\partial \psi}{\partial \nu} \frac{\partial \psi}{\partial \mu} \right) \operatorname{cov}(\hat{\nu}, \hat{\mu}) + \left(\frac{\partial \psi}{\partial \mu} \right)^2 \operatorname{var}(\hat{\mu}). \tag{6.60}$$

An estimate of the asymptotic variance can be obtained by replacing the moments and their variances and covariance by consistent estimates. For the case $\xi = 0$, the MLD index can be written $I_{\text{GE}}^{0} = \log \mu - \nu$ where $\nu = \int \log y \mathrm{d}F(y)$. From (6.60), the asymptotic variance given by the delta method is equal to

[39] From (6.49), (6.50), and (6.57), we have $I_{\text{Atk}}^{\xi}(F) = 1 - [(\xi^2 - \xi) I_{\text{GE}}^{\xi}(F) + 1]^{1/\xi}, \quad \xi \neq 0$ and $I_{\text{Atk}}^{0}(F) = 1 - \exp(-I_{\text{GE}}^{0}(F))$. Variances of Atkinson indices can thus be written as functions of variances of GE indices using the delta method.

$$\frac{1}{\mu^2}\mathrm{var}(\hat{\mu}) - \frac{2}{\mu}\mathrm{cov}(\hat{\mu}, \hat{\nu}) + \mathrm{var}(\hat{\nu}). \tag{6.61}$$

From the IF approach, we have $Z = y/\mu - \log y$, and the asymptotic variance of the MLD index is the variance of Z divided by n,

$$\frac{1}{n}\mathrm{var}(Z) = \frac{1}{n}\left[\frac{1}{\mu^2}\mathrm{var}(y) - \frac{2}{\mu}\mathrm{cov}(y, \log y) + \mathrm{var}(\log y)\right]. \tag{6.62}$$

The two Equations (6.61) and (6.62) are identical, which demonstrates that the delta and the IF methods give the same results. It can be also demonstrated for $\xi \neq 0$.[40]

6.4.3.2 The Mean Deviation and Its Relatives

Now consider the mean deviation, an inequality index that does not belong to the class of decomposable indices (6.28), but does belong to the quasi-additive class (6.30).

$$I_{\mathrm{MD}}(F) := \int |y - \mu(F)| \mathrm{d}F(y).$$

Noting that $I_{\mathrm{MD}}(F)$ can be rewritten as

$$I_{\mathrm{MD}}(F) = 2\int \iota_y [y - \mu(F)] \mathrm{d}F(y) \tag{6.63}$$

where $\iota_y := \iota(y \geq \mu(F))$, the IF is

$$\mathrm{IF}(z; I_{\mathrm{MD}}, F) = 2[\iota_z + \overline{q} - 1][z - \mu(F)] - I_{\mathrm{MD}}(F) \tag{6.64}$$

where $\overline{q} := F(\mu)$. The asymptotic variance of the MD index can be obtained, rewriting the IF as a random variable minus its expectation $\mathrm{IF}(y; I_{\mathrm{MD}}, F) = Z - E(Z)$. From (6.64), we have

$$Z = 2(\overline{q} - 1)y + 2\iota_y[y - \mu(F)] \tag{6.65}$$

From Lemma 1, (6.36), (6.37), the asymptotic variance of $\sqrt{n}\left(I_{\mathrm{MD}}\left(F^{(n)}\right) - I_{\mathrm{MD}}(F)\right)$ is then equal to the variance of Z.

From a sample $(y_1, \ldots, y_n)$, the mean deviation index can be estimated as:

$$\hat{I}_{\mathrm{MD}} := I_{\mathrm{MD}}\left(F^{(n)}\right) = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{\mu}|. \tag{6.66}$$

[40] For extensions to the case of weighted data and complex survey design, see Zheng and Cushing (2001), Cowell and Jenkins (2003), Biewen and Jenkins (2006), and Verma and Betti (2011). For an alternative approach to estimation of Atkinson indices using a Box–Cox transformation, see Guerrero (1987).

The asymptotic variance can be estimated as the empirical variance of $(Z_1, \ldots, Z_n)$ divided by n,

$$\widehat{\mathrm{var}}\left(\hat{I}_{\mathrm{MD}}\right) = \frac{1}{n^2}\sum_{i=1}^{n}(Z_i - \overline{Z})^2, \tag{6.67}$$

where

$$Z_i = 2(\hat{q} - 1)y_i + 2(y_i - \hat{\mu})\iota(y_i \geq \hat{\mu}), \tag{6.68}$$

and

$$\hat{q} := F^{(n)}(\hat{\mu}) = n^{-1}\sum_{i=1}\iota(y_i \leq \hat{\mu})$$

The same methodology with some extra terms can be used to derive the asymptotic variance of the more commonly used relative mean deviation or Pietra ratio

$$\int\left|\frac{y}{\mu(F)} - 1\right|\mathrm{d}F(y).$$

In the literature, the asymptotic variance is usually obtained with the IF method without using it expressed as a function of a random variable minus its expectation. It gives similar numerical results, but formulas and implementation are more complicated. For instance, using Lemma 1 and Equations (6.64) and (6.63), the asymptotic variance of the mean deviation can be derived as follows:

$$\begin{aligned}\int \mathrm{IF}(z; I_{\mathrm{MD}}, F)^2\mathrm{d}F(z) &= 4\int[\iota_z + \overline{q} - 1]^2[z - \mu(F)]^2\mathrm{d}F(z) + I_{\mathrm{MD}}(F)^2 \\ &\quad - 2I_{\mathrm{MD}}(F)\int 2\iota_z[z - \mu(F)]\mathrm{d}F(z) \\ &= 4[\overline{q} - 1]^2\int_{\underline{y}}^{\mu(F)}\left[z - \mu(F)\right]^2 + 4\overline{q}^2\int_{\mu(F)}^{\overline{y}}[z - \mu(F)]^2 - I_{\mathrm{MD}}(F)^2\end{aligned}$$

This formula for the asymptotic variance of the mean deviation index is the same as derived in Gastwirth (1974).

6.4.3.3 The Gini Coefficient

The general form (6.31) is cumbersome, but we can fairly easily derive results for the most important member of this class, namely, the Gini coefficient.

The Gini index can be expressed in a number of different forms. Let us consider the following expressions,

$$I_{\text{Gini}}(F) = \frac{1}{2\mu}\int\int |y - y'| \mathrm{d}F(y)\mathrm{d}F(y'), \tag{6.69}$$

$$= 1 - 2\int_0^1 L(F; q)\mathrm{d}q, \tag{6.70}$$

where $L(F;\ q) = C(F;\ q)/\mu(F)$ is the qth ordinate of the Lorenz curve—see Equation (6.122). Equation (6.69) presents the Gini as the normalized average absolute difference between all the possible pairs of incomes in the population, whereas Equation (6.70) shows that the Gini index is twice the area between the Lorenz curve and the 45° line.

Applying the IF method to the form (6.70), we find the IF of I_{Gini} to be given by Monti (1991):

$$\mathrm{IF}(z; I_{\text{Gini}}, F) = 1 - I_{\text{Gini}}(F) - \frac{2C(F;F(z))}{\mu(F)} + z\frac{1 - I_{\text{Gini}}(F) - 2[1 - F(z)]}{\mu(F)} \tag{6.71}$$

The asymptotic variance of the Gini coefficient has been derived from the IF in Cowell and Victoria-Feser (2003), Bhattacharya (2007), Barrett and Donald (2009), and Davidson (2009a, 2010). A simple formula can be obtained, noting that the IF of the Gini index can be expressed as a random variable minus its expectation, $\mathrm{IF}(z;\ I_{\text{Gini}},\ F) = (Z - E(Z))/\mu(F)$, where[41]

$$Z = [1 - I_{\text{Gini}}(F)]z - 2[C(F;F(z)) + z(1 - F(z))] \tag{6.72}$$

Using Lemma 1, (6.36) and (6.37), one immediately gets the asymptotic variance of $\sqrt{n}\left(I_{\text{Gini}}\left(F^{(n)}\right) - I_{\text{Gini}}(F)\right)$, equal to the variance of Z divided by the square of the mean, $\mathrm{var}(Z)/\mu(F)^2$.

The computation of the Gini index and its variance can be easily obtained in practice. If we define the "positional weight"

$$\kappa(y) := \frac{F(y^-) + F(y^+) - 1}{\mu(F)},$$

where $F(y^-) := \lim_{x\uparrow y} F(x)$ and $F(y^+) := \lim_{x\downarrow y} F(x)$, then the definition (6.69) can alternatively be expressed in the following convenient forms[42]:

$$I_{\text{Gini}}(F) = \int \kappa(y) y \mathrm{d}F(y) \tag{6.73}$$

[41] Note that $E[C(F;\ F(z))] = E[z\ [1 - F(z)]] = [1 - I_{\text{Gini}}(F)]\ \mu(F)/2$.

[42] Using the definition of the Lorenz curve in (6.70), interchanging the order of the integration and simplifying give the result (Davidson, 2009a). For an extensive list of alternative equivalent ways of writing the Gini coefficient—including our expressions (6.69), (6.70), (6.73), and (6.74)—see Yitzhaki and Schechtman (2013).

$$= \frac{2}{\mu} \operatorname{cov}(y, F(y)) \tag{6.74}$$

In other words, the Gini is also equal to the weighted sum of incomes using the κ weights (6.73) and is equal to $2/\mu$ times the covariance between y and $F(y)$ (6.74). For the distribution-free approach, we replace $\mu(F)$ by the sample mean $\hat{\mu}$ and the covariance by an unbiased estimate in (6.74). It leads us to compute the Gini index as:

$$\hat{I}_{\text{Gini}} = I_{\text{Gini}}\left(F^{(n)}\right) = \sum_{i=1}^{n} \kappa_i y_{(i)} \tag{6.75}$$

where the $y_{(i)}$, $i = 1, \ldots, n$, are the order statistics $(y_{(1)} \leq y_{(2)} \leq \cdots \leq y_{(n)})$ and[43]

$$\kappa_i := \frac{2i - n - 1}{\hat{\mu} n (n-1)}.$$

Davidson (2009a) shows that (6.75) is a bias-corrected estimator of the Gini index[44] and proposed estimating the variance of the Gini index as:

$$\widehat{\operatorname{var}}\left(\hat{I}_{\text{Gini}}\right) = \frac{1}{(n\hat{\mu})^2} \sum_{i=1}^{n} (Z_i - \overline{Z})^2 \tag{6.76}$$

where

$$Z_i = -\left(\hat{I}_{\text{Gini}} + 1\right) y_{(i)} + \frac{2i-1}{n} y_{(i)} - \frac{2}{n} \sum_{j=1}^{i} y_{(j)} \tag{6.77}$$

and $\overline{Z} = n^{-1} \sum_{i=1}^{n} Z_i$. Here, the Z_i terms are estimates of the realizations of Z defined in (6.72), where the value of $F(y_{(i)})$ is estimated by $F^{(n)}(y_{(i)}) = (2i-1)/(2n)$.[45] Davidson (2010) extends this approach to derive a variance estimator for the family of S-Gini indices.

[43] Clearly (6.75) is the empirical counterpart of the weighted sum (6.73). Also, using $\operatorname{Cov}(y, F(y)) = E(yF(y)) - E(y)E(F(y))$ in (6.74) and replacing $E(yF(y))$ by $(n-1)^{-1} \sum_{i=1}^{n} y_{(i)}(i/n)$ and $E(y)E(F(y))$ by $\hat{\mu}(n-1)^{-1} \sum_{i=1}^{n} (i/n)$ gives (6.75).

[44] Equation (6.75) is equal to $n/(n-1)$ times Equation (6.5) in Davidson (2009a).

[45] Other estimates of the variance of the Gini index have been proposed in the literature, but they are either complicated or quite unreliable. Nygård and Sandström (1985), Sandström et al. (1988), Cowell (1989), Schechtman (1991), and Bhattacharya (2007) provide formulas that are not easy to implement. A simple method based on OLS regression has been proposed by Ogwang (2000) and Giles (2004), but the standard errors obtained are unreliable, as shown by Modarres and Gastwirth (2006). For a recent review of this literature, see Langel and Tillé (2013) or Yitzhaki and Schechtman (2013); for applications to complex survey designs see Binder and Kovacevic (1995) and Kovacevic and Binder (1997).

6.4.4 Application: Poverty Measures

For a poverty index, we need a poverty line that may be an exogenously given constant ζ or may depend on the income distribution $\zeta(F)$. An important class of poverty indices can then be described in the following way:

$$P(F) := \int p(y, \zeta(F))\mathrm{d}F(y) \tag{6.78}$$

where p is a *poverty evaluation function* that is nonincreasing in y and takes the value zero for $y \geq \zeta(F)$. Once again we need the IF, which is given by

$$\mathrm{IF}(z; P, F) = p(z, \zeta(F)) - P(F) + \int p_\zeta(y, \zeta)\mathrm{d}F(y)\mathrm{IF}(z; \zeta, F) \tag{6.79}$$

where p_ζ is the differential of p with respect to its second argument (Cowell and Victoria-Feser, 1996a). It is clear from (6.79) that the form for the asymptotic variance of the poverty index will depend on the precise way in which the poverty line depends on the income distribution. The following specifications cover almost all the versions encountered in practice

$$\zeta(F) = \zeta_0 + \gamma\mu(F), \tag{6.80}$$

or

$$\zeta(F) = \zeta_0 + \gamma y_q, \quad q \in \mathbb{Q}, \tag{6.81}$$

where y_q is defined in (6.24). The interpretation is that the poverty line could be tied to the mean, as in (6.80), in which case we have

$$\mathrm{IF}(z; \zeta, F) = \gamma \mathrm{IF}(z; \mu, F) = \gamma[z - \mu(F)] \tag{6.82}$$

or to a quantile (6.81), such as the median, in which case we have

$$\mathrm{IF}(z; \zeta, F) = \gamma \frac{q - \iota(y_q \geq z)}{f(y_q)}. \tag{6.83}$$

The asymptotic variance can be immediately calculated from Equations (6.79) and (6.82) or (6.83). Let us take the simple case where $\gamma = 0$[46] so that one has the exogenous poverty line ζ_0. Equation (6.79) yields the IF $p(z, \zeta_0) - P(F)$ and so, using Lemma 1, we find the asymptotic variance of $P(F)$ in (6.78) to be

$$\int p(z, \zeta_0)^2 \mathrm{d}F(z) - P(F)^2.$$

[46] Note that if $\gamma > 0$ then to estimate the asymptotic variance of P using (6.82) one needs information on the whole distribution; with (6.83) one needs a density estimate at y_q.

The asymptotic variance of the poverty index is then equal to the variance of the poverty evaluation function, var($p(y, \zeta_0)$). We can see that the preceding IF is expressed as a function of a random variable minus its expectation,

$$\mathrm{IF}(y; P, F) = Z - E(Z) \quad \text{where } Z = p(y, \zeta_0) \tag{6.84}$$

From (6.36) and (6.37), the asymptotic variance is the variance of Z.

A second important class of poverty indices consists of those in the *rank-dependent* form—compare (6.31) above—and can be described in the following way:

$$P_{\mathrm{RD}}(F) := \int p(y, \zeta(F), F(y))\mathrm{d}F(y) \tag{6.85}$$

Comparing (6.85) with (6.78), we see that the poverty evaluation function p has an extra argument reflecting the individual's rank in the population. The IF for this class of poverty measures is more complicated (Cowell and Victoria-Feser, 1996a), and we deal with this separately in Sections 6.4.4.2 and 6.4.4.3.

6.4.4.1 Foster–Greer–Thorbecke (FGT)

For a fixed poverty line ζ_0 the widely used class of poverty indices introduced by Foster et al. (1984) belongs to the class (6.78) and has the form

$$P^{\xi}_{\mathrm{FGT}}(F) = \int_0^{\zeta_0} \left(\frac{\zeta_0 - y}{\zeta_0}\right)^{\xi} \mathrm{d}F(y) \quad \xi \geq 0, \tag{6.86}$$

When $\xi = 0$, the FGT poverty measure is equal to the headcount ratio, which gives the proportion of individuals living in poverty, $F(\zeta_0)$. This index is insensitive to the distribution of incomes among the poor and, therefore, to the depth of poverty. When $\xi = 1$, the FGT poverty measure is the poverty gap index, which considers how far, on the average, the poor are from that poverty line. This index captures the depth of poverty, but it is insensitive to some types of transfers among the poor and, therefore, to some distributional aspects of poverty. Let y_i, $i = 1, \ldots, n$, be an IID sample from the distribution F. The FGT poverty indices (6.86) can be estimated consistently as follows:

$$\hat{P}^{\xi}_{\mathrm{FGT}} := P^{\xi}_{\mathrm{FGT}}\left(F^{(n)}\right) = \frac{1}{n}\sum_{i=1}^{n_p} \left(\frac{\zeta_0 - y_{(i)}}{\zeta_0}\right)^{\xi}, \tag{6.87}$$

where n_p is the number of individuals with incomes not greater than the poverty line, that is, the number of poor, and $y_{(i)}$, $i = 1, \ldots, n$ are the order statistics $(y_{(1)} \leq y_{(2)} \leq \cdots \leq y_{(y)})$. This estimate is asymptotically Normal, with a variance that can be estimated as:[47]

[47] See Kakwani (1993).

$$\widehat{\text{var}}\left(\hat{P}^{\xi}_{\text{FGT}}\right) = \frac{1}{n}\left(\hat{P}^{2\xi}_{\text{FGT}} - \left(\hat{P}^{\xi}_{\text{FGT}}\right)^2\right). \tag{6.88}$$

We can also estimate the asymptotic variance of the FGT index using the IF expressed as a function of a random variable minus its expectation. From (6.78), (6.84), and (6.86), we have $\text{IF}(y, P^{\xi}_{\text{FGT}}, F) = Z - E(Z)$, where

$$Z_i = \left|1 - \frac{y}{\zeta_0}\right|^{\xi} \iota(y \le \zeta_0) \tag{6.89}$$

From Lemma 1, (6.36), and (6.37), the asymptotic variance of $\sqrt{n}\left(P_{\text{FGT}}\left(F^{(n)}\right) - P_{\text{FGT}}(F)\right)$ is then equal to the variance of Z. From a sample $(y_1, \ldots, y_n)$, let us define

$$Z_i = \begin{cases} \left[\left(\zeta_0 - y_{(i)}\right)/\zeta_0\right]^{\xi} & \text{for } i \le n_p \\ 0 & \text{for } i > n_p \end{cases} \tag{6.90}$$

where the $y_{(i)}$, $i = 1, \ldots, n$ are the order statistics. We can see that

$$\hat{P}^{\xi}_{\text{FGT}} = \frac{1}{n}\sum_{i=1}^{n} Z_i = \overline{Z} \quad \text{and} \quad \widehat{\text{var}}\left(\hat{P}^{\xi}_{\text{FGT}}\right) = \frac{1}{n^2}\sum_{i=1}^{n}(Z_i - \overline{Z}). \tag{6.91}$$

The FGT index can be estimated by the mean of Z_i, for $i = 1, \ldots, n$, with a variance estimated by the empirical variance of Z_i, divided by n. The two approaches, in Equations (6.87), (6.88), and (6.91), give similar numerical results.[48]

6.4.4.2 Sen Poverty Index

The Sen poverty index (Sen, 1976) belongs to the class (6.85) and can be expressed as the average of the headcount ratio and the poverty gap index, weighted by the Gini coefficient among the poor,

$$P_{\text{Sen}}(F) = P^0_{\text{FGT}} I^p_{\text{Gini}} + P^1_{\text{FGT}}(1 - I^p_{\text{Gini}}), \tag{6.92}$$

where I^p_{Gini} is the Gini index computed with incomes below the poverty line. When the distribution of incomes among the poor is equal, $I^p_{\text{Gini}} = 0$, the Sen index is equal to the headcount ratio ($P_{\text{Sen}} = P^0_{\text{FGT}}$). When the distribution of incomes among the poor is extremely unequal, $I^p_{\text{Gini}} = 1$, the Sen index is equal to the poverty gap index ($P_{\text{Sen}} = P^1_{\text{FGT}}$).The Sen poverty measure can be written

$$P_{\text{Sen}}(F) = \frac{2}{\zeta_0 F(\zeta_0)}\int_0^{\zeta_0} (\zeta_0 - y)(F(\zeta_0) - F(y))\mathrm{d}F(y) \tag{6.93}$$

[48] The problem of estimation in the presence of complex survey design is addressed in Howes and Lanjouw (1998), Zheng (2001), Berger and Skinner (2003), and Verma and Betti (2011).

From Davidson (2009a), we can derive the IF as a function of a random variable minus its expectation, $\mathrm{IF}(z, P_{\mathrm{Sen}}, F) = \frac{2}{\zeta_0 F(\zeta_0)}(Z - E(Z))$, where[49]

$$Z = \left[\zeta_0 F(\zeta_0) - \frac{\zeta_0 P_S}{2} - zF(\zeta_0) + zF(z) - C(F;F(z))\right]\iota(z \le \zeta_0) \tag{6.94}$$

A consistent estimate of the Sen poverty index can be obtained by replacing F by $F^{(n)}$ in (6.93),[50]

$$\hat{P}_{\mathrm{Sen}} := P_{\mathrm{Sen}}\left(F^{(n)}\right) = \frac{2}{nn_p\zeta_0}\sum_{i=1}^{n_p}\left(\zeta_0 - y_{(i)}\right)\left(n_p - i + \frac{1}{2}\right), \tag{6.95}$$

where the value of $F(y_{(i)})$ is estimated by $F^{(n)}(y_{(i)}) = (2i-1)/(2n)$. This estimate is asymptotically Normal, with a variance that can be computed as follows:

$$\widehat{\mathrm{var}}\left(\hat{P}_{\mathrm{Sen}}\right) = \frac{4}{\left(\zeta_0 n_p\right)^2}\sum_{i=1}^{n}(Z_i - \overline{Z})^2 \tag{6.96}$$

where

$$Z_i = \frac{\zeta_0}{2}\left(\frac{2n_p}{n} - \hat{P}_{\mathrm{Sen}}\right) - \frac{2n_p - 2i + 1}{2n}y_{(i)} - \frac{1}{n}\sum_{j=1}^{i} y_{(j)} \tag{6.97}$$

for $i = 1, \ldots, n_p$, and $Z_i = 0$ for $i = n_p + 1, \ldots, n$, with $\overline{Z} = n^{-1}\sum_{i=1}^{n} Z_i$. Here, Z_i are estimates of the realizations of Z defined in (6.94).[51]

6.4.4.3 Sen–Shorrocks–Thon Poverty Index

The Sen–Shorrocks–Thon (SST) index is a convenient modified version of the Sen poverty index, defined as follows,

$$P_{\mathrm{SST}}(F) = P^0_{\mathrm{FGT}} P^1_{\mathrm{FGT}_g}\left(1 + I^{pg}_{\mathrm{Gini}}\right) \tag{6.98}$$

where $P^1_{\mathrm{FGT}_g}$ is the poverty gap index computed with incomes below the poverty line, and I^{pg}_{Gini} is the Gini coefficient computed with individuals' poverty gap ratios rather than individuals' incomes for the whole population $((\zeta_0 - y_i)/\zeta_0$ rather than y_i,

[49] In Equation (6.50) in Davidson (2009a), replacing y_i in the expression of the summand by z yields the IF. The relationship to the IF is not related in the last paper; it is related in Davidson (2010), where the same method is used with S-Gini indices.

[50] This expression does not coincide exactly with Sen's own definition for a discrete population. See Appendix A in Davidson (2009a) for a discussion of this point.

[51] Another variance estimator has been proposed by Bishop et al. (1997).

for $i = 1, \ldots, n$).[52] The Gini coefficient of poverty gap ratios can be viewed as a measure of poverty inequality in a society. The SST index satisfies the transfer and continuity axioms, whereas the Sen index does not.[53]

This index can be decomposed into

$$\Delta \log P_{\mathrm{SST}} = \Delta \log P^0_{\mathrm{FGT}} + \Delta P^1_{\mathrm{FGT}_g} + \Delta \log\left(1 + I^{pg}_{\mathrm{Gini}}\right) \tag{6.99}$$

A percentage change in SST can then be viewed as the sum of percentage changes in the proportion of poor, the average poverty gap among the poor, and one plus the Gini index of poverty gaps for the population. The poverty is decomposed in three aspects: Are there more poor? Are the poor poorer? Is there higher poverty inequality in the society?

The SST poverty index can be written

$$P_{\mathrm{SST}}(F) = \frac{2}{\zeta_0}\int_0^{\zeta_0} (\zeta_0 - y)(1 - F(y))\mathrm{d}F(y) \tag{6.100}$$

As in Section 6.4.4.2, we derive the IF as a function of a random variable minus its expectation, $\mathrm{IF}(z, P_{\mathrm{SST}}, F) = \frac{2}{\zeta_0}(Z - E(Z))$, where[54]

$$Z = [\zeta_0(1 - F(\zeta_0)) - z(1 - F(z)) + C(F;F(\zeta_0)) - C(F;F(z))]\iota(z \le \zeta_0) \tag{6.101}$$

The SST poverty index can be consistently estimated as:[55]

$$\hat{P}_{\mathrm{SST}} := P_{\mathrm{SST}}\left(F^{(n)}\right) = \frac{2}{\zeta_0 n(n-1)}\sum_{i=1}^{n_p}\left(\zeta_0 - y_{(i)}\right)(n - i) \tag{6.102}$$

It is asymptotically Normal, with an estimator of the variance given by

$$\widehat{\mathrm{var}}\left(\hat{P}_{\mathrm{SST}}\right) = \frac{4}{\zeta_0^2 (n-1)^2}\sum_{i=1}^{n}(Z_i - \overline{Z})^2 \tag{6.103}$$

where

[52] The original index was proposed by Shorrocks (1995). Xu and Osberg (2002) show that it can be written as Equation (6.98). They also show that the Sen index is equal to $S = P^0_{\mathrm{FGT}} P^1_{\mathrm{FGT}_g}(1 + I^{pg}_{\mathrm{Gini}})$ and therefore that the SST index differs from the Sen index because it uses the Gini index of poverty gap ratios for the whole population, whereas the Sen index uses the Gini index of poverty gap ratios for the poor.

[53] The (strong upward) transfer axiom states that an increase in a poverty measure should occur if the poorer of the two individuals involved in an upward transfer of income is poor, and even if the beneficiary crosses the poverty line.

[54] In the first equation of column two in Davidson (2009a, p. 39), replacing y_i in the expression of the summand by z yields the IF. The relationship to the IF is not related in the last paper; it is related in Davidson (2010), where the same method is used with S-Gini indices.

[55] The SST index defined by Shorrocks (1995) is obtained by replacing F by $\hat{F}$ in (6.100). Here, (6.102) is a simplified version of the bias-corrected estimator in the last equation in Davidson (2009a, p. 37). In the case of complex survey data, see Osberg and Xu (2000).

$$Z_i = \zeta_0\left(1-\frac{n_p}{n}\right) - \frac{2n-2i+1}{2n}y_{(i)} + \frac{1}{n}\sum_{j=1}^{n_p} y_{(j)} - \frac{1}{n}\sum_{j=1}^{i} y_{(j)} \tag{6.104}$$

for $i=1, \ldots, n_p$, and $Z_i=0$ for $i=n_p+1, \ldots, n$, with $\overline{Z}=n^{-1}\sum_{i=1}^{n} Z_i$. Here, Z_i are estimates of the realizations of Z defined in (6.101).

6.4.5 Finite Sample Properties

6.4.5.1 Asymptotic and Bootstrap Methods

Asymptotic Normality allows us to perform asymptotic inference. In practice, we are concerned with finite samples, and asymptotic inference can be unreliable. When asymptotic inference does not perform well in a finite sample, bootstrap methods can be used to perform accurate inference. The bootstrap appears to be an ideal method for inference with inequality and poverty indices because the observations of the sample are often IID.

Let us consider a welfare index W and its sample counterpart $\hat{W}$. An asymptotic confidence interval at 95% would be computed as

$$\mathrm{CI}_{\mathrm{asym}} = \left[\hat{W} - c_{0.975}\widehat{\mathrm{var}}\left(\hat{W}\right)^{1/2}; \hat{W} - c_{0.025}\widehat{\mathrm{var}}\left(\hat{W}\right)^{1/2}\right] \tag{6.105}$$

where $c_{0.025}$ and $c_{0.975}$ are the 2.5th and 97.5th percentiles of the asymptotic distribution of the t-statistic, $t=\left(\hat{W}-W_0\right)/\widehat{\mathrm{var}}\left(\hat{W}\right)$, where W_0 is the true value of the welfare index. In general, the asymptotic distribution of the t-statistic is the standard Normal distribution, from which we have $c_{0.975}=-c_{0.025}\approx 1.96$.

When the bootstrap is used in combination with the asymptotic variance estimate, asymptotic refinements can be obtained over asymptotic methods.[56] To compute a bootstrap confidence interval, we generate B samples of size n by resampling with replacement from the observed sample. For bootstrap sample b, we compute the index $\hat{W}_b^*$, its variance estimate $\widehat{\mathrm{var}}\left(\hat{W}_b^*\right)$, and the t-statistic $t_b^*=\left(\hat{W}_b^*-\hat{W}\right)/\widehat{\mathrm{var}}\left(\hat{W}_b^*\right)^{1/2}$. A bootstrap confidence interval at 95% would be computed as

$$\mathrm{CI}_{\mathrm{boot}} = \left[\hat{W} - c^*_{0.975}\widehat{\mathrm{var}}\left(\hat{W}\right)^{1/2}; \hat{W} - c^*_{0.025}\widehat{\mathrm{var}}\left(\hat{W}\right)^{1/2}\right] \tag{6.106}$$

where $c^*_{0.025}$ and $c^*_{0.975}$ are the 2.5th and 97.5th percentiles of the EDF of the bootstrap t-statistics, that is, the $\lceil 0.025B\rceil$ and $\lceil 0.975B\rceil$ order statistics of the t_b^*, where $\lceil x\rceil$ denotes the smallest integer not smaller than x. In this approach, the unknown distribution of the population is replaced by the EDF of the original sample, from which we generate bootstrap samples and compute t-statistics testing the (true) hypothesis that the index is equal

[56] See Beran (1988). It means that the bootstrap method presented in this section provides an asymptotic refinement over the percentile bootstrap proposed in Mills and Zandvakili (1997).

to $\hat{W}$. The simulated distribution of the bootstrap t-statistics is used, as an approximation of the unknown distribution of t, to calculate critical values.

The bootstrap can also be used to test hypotheses and to compute p-values. To test the hypothesis that the population value of the index is W_0, for a one-tailed test, the bootstrap p-value would be the proportion of the t_b^* that are more extreme than the t-statistic computed from the observed sample t. Here, the bootstrap test is also based on the EDF of the bootstrap t-statistics t_b^*. The null hypothesis is rejected at a significance level 0.05 if the bootstrap p-value is less than 0.05. To test the hypothesis that two indices are the same from two populations, a suitable t-statistic is $\tau = \left(\hat{W}_1 - \hat{W}_2\right)/\left(\widehat{\text{var}}\left(\hat{W}_1\right) + \widehat{\text{var}}\left(\hat{W}_2\right)\right)^{1/2}$ when the samples are independent. For bootstrap sample b, the bootstrap statistic is computed then as $\tau_b^* = \left(\hat{W}_{1b}^* - \hat{W}_{2b}^* - \hat{W}_1 + \hat{W}_2\right)/\left(\widehat{\text{var}}\left(\hat{W}_{1b}^*\right) + \widehat{\text{var}}\left(\hat{W}_{2b}^*\right)\right)^{1/2}$. When the samples are dependent, the statistic should take account of the covariance, and the bootstrap samples should be generated by resampling pairs of observations with replacement. Again, the bootstrap p-value would be the proportion of the τ_b^* that is more extreme than τ.

6.4.5.2 Simulation Evidence

We now turn to the performance in a finite sample of inference based on inequality and poverty measures. The coverage rate of a confidence interval is the probability that the random interval does include, or cover, the true value of the parameter. A method of constructing confidence intervals with good finite sample properties should provide a coverage rate close to the nominal confidence level. For a confidence interval at 95%, the nominal coverage rate is equal to 95%. In this section, we use Monte Carlo simulation to approximate the coverage rate of asymptotic and bootstrap confidence intervals in several experimental designs.

In our experiments, data are generated from lognormal distributions, $\Lambda(y;\, 0,\, \sigma)$, and from Singh–Maddala distributions, $\text{SM}(y;\, 2.8,\, 0.193,\, q)$. As σ increases and q decreases the upper tail of the distribution decays more slowly. The sample size is $n=500$, the number of bootstrap samples is $B=499$, and the number of experiments is $N=10{,}000$.[57] When poverty indices are used, the poverty line is computed as half the median.

Table 6.6 presents coverage of asymptotic and bootstrap confidence intervals at the 95% level for the Theil, MLD, Gini, and SST indices. The results show that asymptotic and bootstrap confidence intervals are reliable when we consider the SST poverty index. Indeed, the coverage rates of the SST index are always close to the nominal coverage rate of 95%. In contrast, when we consider inequality measures, bootstrap confidence intervals outperform asymptotic confidence intervals, but they become less reliable as σ increases and q decreases. In other words, asymptotic and bootstrap inferences deteriorate

[57] For well-known reasons—see Davison and Hinkley (1997) or Davidson and MacKinnon (2000)—the number of bootstrap resamples B should be chosen so that $(B+1)/100$ is an integer.

Table 6.6 Coverage of asymptotic and bootstrap confidence intervals at the 95% level for the Theil, MLD, Gini, and SST indices, $n=500$

	Theil		MLD		Gini		SST	
	asym	**boot**	**asym**	**boot**	**asym**	**boot**	**asym**	**boot**
Lognormal								
$\sigma=0.5$	0.927	0.936	0.936	0.942	0.942	0.943	0.926	0.952
$\sigma=1.0$	0.871	0.913	0.922	0.936	0.922	0.936	0.945	0.940
$\sigma=1.5$	0.746	0.854	0.888	0.921	0.876	0.920	0.964	0.937
Singh–Maddala								
$q=1.7$	0.915	0.931	0.938	0.945	0.945	0.944	0.945	0.950
$q=1.2$	0.856	0.905	0.913	0.930	0.925	0.934	0.945	0.951
$q=0.7$	0.647	0.802	0.820	0.890	0.847	0.906	0.939	0.946

as the upper tail of the underlying distribution is heavier. For instance, asymptotic confidence intervals cover the true value of the Theil index 64.7% of times when the underlying distribution is the Singh–Maddala with $q=0.7$. Bootstrap confidence intervals provide better results, with a coverage rate of 80.2%, but it is still significantly different from the expected 95%. Note that the Theil index is known to be more sensitive to the upper tail of the distribution than the MLD and Gini, and confidence intervals with the Theil index are slightly less reliable than with the MLD and Gini indices.

These results illustrate that asymptotic and bootstrap inference on inequality measures is sensitive to the exact nature of the upper tail of the income distribution. Bootstrap inference on inequality measures are expected to perform reasonably well in moderate and large samples, unless the tails are quite heavy.[58] Moreover, asymptotic and bootstrap inference on poverty measures perform well in finite samples.

6.4.5.3 Inference with Heavy-Tailed Distributions

When the distribution is one with quite a heavy upper tail, asymptotic and bootstrap inferences are known to perform poorly in finite samples. Several approaches have been proposed in the literature to obtain more reliable inference.

Schluter and van Garderen (2009) and Schluter (2012) proposed normalizing transformation of the index, before using the bootstrap, in order to use a statistic with a distribution closer to the Normal. Let g denote a transformation of the index W; a standard bootstrap confidence interval can be obtained on the transformed index $g(W)$ and, therefore, on the untransformed index by inverting the relation between the welfare index and the parameters. Let $c^*_{0.025}$ and $c^*_{0.975}$ be the 2.5th and 97.5th percentiles of the EDF of the bootstrap t-statistics

[58] Additional results with other distributions, other indices, and hypotheses testing can be found in Davidson and Flachaire (2007), Cowell and Flachaire (2007), and Davidson (2009a, 2010, 2012).

$$t_b^* = \frac{g(\hat{W}_b^*) - g(\hat{W})}{g'(\hat{W}_b^*)\widehat{\text{var}}(\hat{W}_b^*)^{1/2}},$$

where g' is the first derivative of g, a bootstrap confidence interval at 95% for W would then be defined as

$$\left[g^{-1}\left(g(\hat{W}) - c_{0.025}^* g'(\hat{W})\widehat{\text{var}}(\hat{W})^{1/2}\right); g^{-1}\left(g(\hat{W}) - c_{0.975}^* g'(\hat{W})\widehat{\text{var}}(\hat{W})^{1/2}\right)\right],$$

if g^{-1} is nondecreasing, otherwise $c_{0.025}^*$ and $c_{0.975}^*$ should be interchanged. For instance, Schluter (2012) exploits a systematic relationship between the inequality estimate and its estimated variance to propose variance stabilizing transforms of the index. He suggests computing confidence intervals based on the following transform of the index,

$$g(W) = -\frac{2}{\gamma_2}\exp\left(-\frac{\gamma_1}{2} - \frac{\gamma_2}{2}W\right), \tag{6.107}$$

where γ_1 and γ_2 are the intercept and the slope of a (systematic) linear relation between the index $\hat{W}$ and the logarithmic transformation of its variance estimates $\widehat{\text{var}}(\hat{W})$, and $\gamma_2 > 0$. The parameters γ_1 and γ_2 are estimated by OLS estimation from the regression

$$\log \widehat{\text{var}}(\hat{W}) = \gamma_1 + \gamma_2 \hat{W} + \varepsilon,$$

where realizations of $\widehat{\text{var}}(\hat{W})$ and $\hat{W}$ are obtained by a preliminary bootstrap. For the specific transform (6.107), the inverse function is equal to

$$g^{-1}(x) = -\frac{2}{\gamma_2}\log\left(-\frac{\gamma_2}{2}x\right) - \frac{\gamma_1}{\gamma_2} \tag{6.108}$$

A bootstrap confidence interval at 95% can be computed using (6.107) and (6.108) in the confidence interval defined earlier.

Davidson and Flachaire (2007) and Cowell and Flachaire (2007) considered a semi-parametric bootstrap, where bootstrap samples are generated from a distribution that combines a parametric estimate of the upper tail with a nonparametric estimate of the rest of the distribution. The upper tail is modeled by a Pareto distribution with parameter α estimated by the Hill estimator on the k greatest order statistics of a sample of size n, for some integer $k \leq n$,

$$\hat{\alpha} := \left(\frac{1}{k}\sum_{i=0}^{k-1} \log y_{(n-i)} - \log y_{n-k+1}\right)^{-1} \tag{6.109}$$

where $y_{(j)}$ is the jth order statistic of the sample. Each observation of a bootstrap sample is, with probability p_{tail}, a drawing from the CDF of the Pareto distribution $F(y) = 1 - (y/y_0)^{-\hat{\alpha}}, y > y_0$, where y_0 is the order statistic of rank $n(1 - p_{\text{tail}})$, and, with probability $1 - p_{\text{tail}}$, a drawing from the empirical distribution of the sample of smallest

$n(1-p_{\text{tail}})$-order statistics. For the bootstrap to test a true null hypothesis, we need to compute the value of the welfare index for the bootstrap distribution defined earlier. The CDF of the bootstrap distribution can be written as

$$F_s(y) = \frac{1}{n}\sum_{i=1}^{n(1-p_{\text{tail}})} \iota\left[y_{(i)} \le y\right] + \iota[y \ge y_0]p_{\text{tail}}\left(1-(y/y_0)^{-\hat{\alpha}}\right), \tag{6.110}$$

where $\iota(\cdot)$ is the indicator function (6.1). Indices of interests are functionals of the income distribution and so the index for this bootstrap distribution, $\hat{W}_s$, can be computed.[59] A bootstrap confidence interval can be computed as defined in (6.106), where $c^*_{0.025}$ and $c^*_{0.975}$ are the 2.5th and 97.5th percentiles of the EDF of the bootstrap t-statistics $t^*_b = \left(\hat{W}^*_b - \hat{W}_s\right)/\widehat{\text{var}}\left(\hat{W}^*_b\right)^{1/2}$ In practice, k and p_{tail} are chosen *a priori*. The number of observations k used to compute the Hill estimator can be selected such that $\hat{\alpha}$ does not vary significantly when more observations are taken, and p_{tail} can be chosen such that the resampling from the Pareto distribution is based on a smaller proportion of observation than k/n. It leads the previous authors to select $k = n^{1/2}$ and $p_{\text{tail}} = hk/n$ with $0 < h \le 1$ in their experiments.

An alternative approach could be to generate bootstrap samples from a distribution estimated by finite mixture models. It allows us to estimate any density function, by allowing the number of components to vary, and, once the number of component is selected, to use a parametric distribution to generate bootstrap samples (see Section 6.3.3). For the bootstrap to test a true null hypothesis, we need to compute the value of the welfare index for the mixture distribution, $\hat{W}_m$. With additively decomposable inequality measures, the index for the mixture distribution is easy to calculate because the mixture distribution is a decomposition by groups. For instance, the class of GE indices can be expressed as a simple additive function of within-group and between-group inequality. Let there be K groups, and let the proportion of the population falling in group k be p_k, the class of GE indices is equal to[60]

$$I^{\xi}_{\text{GE}} = \sum_{k=1}^{K} p_k\left[\frac{\overline{y}_k}{\overline{y}}\right]^{\xi} I^{\xi}_{\text{GE},k} - \frac{1}{\xi^2-\xi}\left(\sum_{k=1}^{K} p_k\left[\frac{\overline{y}_k}{\overline{y}}\right]^{\xi} - 1\right) \tag{6.111}$$

where $\overline{y}_k$ is the mean income in group k, $\overline{y}$ is the mean income of the population $\left(\overline{y} = K^{-1}\sum_{k=1}^{K} p_k\overline{y}_k\right)$, and $I^{\xi}_{\text{GE},k}$ is the GE index in group k. For an income distribution estimated as a finite mixture of lognormal distributions,

[59] For instance, the Theil index would be equal to $\hat{I}_s = \nu_s/\mu_s - \log\mu_s$, where $\mu_s = n^{-1}\sum_{i=1}^{n(1-p_{\text{tail}})} y_{(i)} + p_{\text{tail}}\hat{\alpha}y_0/(\hat{\alpha}-1)$ and $\nu_s = n^{-1}\sum_{i=1}^{n(1-p_{\text{tail}})} y_{(i)}\log y_{(i)} + p_{\text{tail}}[\log y_0 + 1/(\hat{\alpha}-1)]\hat{\alpha}y_0/(\hat{\alpha}-1)$.

[60] See Cowell (2011).

$$F_m(y) = \sum_{k=1}^{K} \hat{\pi}_k \Lambda(y; \hat{\mu}_k, \hat{\sigma}_k), \tag{6.112}$$

the value of the GE index is then equal to (6.111) with $p_k = \hat{\pi}_k$ and $\overline{y}_k = \exp\left(\hat{\mu}_k + \hat{\sigma}_k^2/2\right)$. The Gini index is not additively decomposable, but a formula can be found in Young (2011) for a mixture of lognormal distributions. Bootstrap samples are generated from the mixture distribution $F_m(y)$, and a bootstrap confidence interval is computed as defined in (6.106), where $c^*_{0.025}$ and $c^*_{0.975}$ are the 2.5th and 97.5th percentiles of the EDF of the bootstrap t-statistics $t^*_b = \left(\hat{W}^*_b - \hat{W}_s\right)/\widehat{\text{var}}\left(\hat{W}^*_b\right)^{1/2}$.

Table 6.7 presents coverage of asymptotic and bootstrap confidence intervals at the 95% level for the Theil index, with $n = 500$. The first two columns correspond to asymptotic (asym) and standard bootstrap (boot) methods; they reproduce the results given in Table 6.6, given here as benchmarks. The other columns show the results for the alternative bootstrap methods presented earlier. Results obtained by the approach proposed by Schluter (2012) are presented in the third column (varstab), bootstrapping a variance stabilizing transform of the Theil index. In the fourth column, the semiparametric bootstrap proposed by Davidson and Flachaire (2007) and Cowell and Flachaire (2007) is used to generate bootstrap samples (semip), with $k = n^{1/2}$ and $h = 0.6$. Finally, bootstrap samples generated from a mixture of lognormal distributions are considered in the last column (mixture). The simulation results show that, in the presence of very heavy-tailed distributions ($\sigma = 1.5$, $q = 0.7$), significant improvements can be obtained with alternative methods over asymptotic and standard bootstrap methods. However, none of the alternative methods provides very good results overall.

6.4.5.4 *Testing Equality of Inequality Measures*

Confidence intervals are often used to make comparisons between two or more samples. The values of an index computed from independent samples are statistically different if

Table 6.7 Coverage of asymptotic and bootstrap confidence intervals at the 95% level for the Theil index, for several bootstrap approaches, $n = 500$

	asym	boot	varstab	semip	mixture
Lognormal					
$\sigma = 0.5$	0.927	0.936	0.939	0.937	0.942
$\sigma = 1.0$	0.871	0.913	0.907	0.921	0.946
$\sigma = 1.5$	0.746	0.854	0.850	0.915	0.944
Singh–Maddala					
$q = 1.7$	0.915	0.931	0.933	0.926	0.928
$q = 1.2$	0.856	0.905	0.899	0.905	0.912
$q = 0.7$	0.647	0.802	0.796	0.871	0.789

the confidence intervals do not intersect. We can thus test if inequality or poverty measures are different between several countries, or over different periods of time, by comparing their confidence intervals. However, the previous results suggest that this approach may be unreliable when comparing inequality measures if the underlying distributions are quite heavy tailed.

Another principal way of performing inference is by carrying out hypothesis tests. Testing the equality of inequality measures with a t-statistic, Dufour et al. (2013) showed that almost exact inference can be obtained with permutation tests, if the samples come from distributions not too far away from each other, even with very heavy-tailed distributions and very small samples. They also showed that this method outperforms other methods when the distributions are far away from each other.

Let us consider two independent samples $X = \{x_1, x_2, \ldots, x_n\}$ and $Y = \{y_1, y_2, \ldots, y_m\}$ assumed to be two sets of n and m independent observations from distributions F_x and F_y. The null hypothesis that an inequality measure W is the same in the two distributions, H_0: $W_x = W_y$, can be tested with a t-statistic,

$$\tau = \left[\hat{W}_x - \hat{W}_y\right] / \left[\widehat{\text{var}}\left(\hat{W}_x\right) + \widehat{\text{var}}\left(\hat{W}_y\right)\right]^{1/2}, \tag{6.113}$$

where τ follows asymptotically the standard Normal distribution. A standard bootstrap approach would be to generate bootstrap samples X^* and Y^* by resampling with replacement, respectively, n observations from X and m observations from Y. The bootstrap samples are drawings from X and Y, from which an inequality measure would provide different numerical results. The null hypothesis tested from the original sample is then not respected in the bootstrap data-generating process. A modified t-statistic has to be computed to test a true null hypothesis from a bootstrap sample (X_b^*, Y_b^*),

$$\tau_b^* = \left[\hat{W}_{x_b^*} - \hat{W}_{y_b^*} - \left(\hat{W}_x - \hat{W}_y\right)\right] / \left[\widehat{\text{var}}\left(\hat{W}_{x_b^*}\right) + \widehat{\text{var}}\left(\hat{W}_{y_b^*}\right)\right]^{1/2}, \tag{6.114}$$

The bootstrap distribution is the EDF of the B bootstrap statistics, τ_b^* for $b = 1, \ldots, B$; it is used as an approximation of the true distribution of the t-statistic τ.[61]

In the permutation approach, we generate samples by permuting the combined sample of N observations,

$$(X, Y) = \{x_1, \ldots, x_n, y_1, \ldots, y_m\}, \tag{6.115}$$

where $N = n + m$. The permutation samples X^* and Y^* are composed, respectively, by the first n and the remaining m observations of the permuted combined sample. Note that the combined sample can be permuted by resampling *without* replacement N observations from (X, Y).

[61] Davidson and Flachaire (2007) considered testing the difference of two inequality measures with independent samples, but no significant improvement of their semiparametric bootstrap method over standard bootstrap method is found.

The permutation samples are drawings from the same set of observations; the null hypothesis tested from the original sample is then respected in the data-generating process. From a permutation sample (X_p^*, Y_p^*), the permutation *t*-statistic is

$$\tau_p^* = \left[\hat{W}_{x_p^*} - \hat{W}_{y_p^*}\right] / \left[\widehat{\text{var}}\left(\hat{W}_{x_p^*}\right) + \widehat{\text{var}}\left(\hat{W}_{y_p^*}\right)\right]^{1/2}. \qquad (6.116)$$

The permutation distribution is the EDF of the P permutation statistics, τ_p^* for $p = 1, \ldots, P$; again it is used as an approximation of the true distribution of the *t*-statistic τ. If the underlying distributions are identical, permutation tests are known to provide *exact* inference in finite samples.[62] Chung and Romano (2013) show that under weak assumptions for comparing estimators, permutation tests are asymptotically valid when $F_x \neq F_y$, while retaining the exact rejection probability in finite samples when $F_x = F_y$.[63] Testing the equality of inequality measures, Dufour et al. (2013) showed that almost exact inference is obtained with permutation tests in very small samples with heavy-tailed distributions, if the samples come from distributions not too far away from each other.

To illustrate, let us consider a simulation experiment concerned with testing the equality of the Gini index, H_0: $I_{\text{Gini}}(F_x) = I_{\text{Gini}}(F_y)$. Data are generated from Singh–Maddala distributions, SM(y; a, b, q), with parameters chosen such that the Gini index is identical in all distributions. The upper tail of a Singh–Maddala distribution behaves like a Pareto distribution with shape parameter $\alpha = aq$.[64] So, the smaller is α, the heavier is the upper tail. The sample size is very small, $n = m = 50$, and the distributions can be very heavy tailed to stress-test the methods employed in testing. The number of replications is equal to 10,000, and the number of bootstrap and permutation samples are $B = P = 999$. We compute the rejection probability, or rejection frequency, as the proportion of p-values less than a nominal level equal to 0.05.[65] Inference is exact if the rejection probability is equal to 0.05.

[62] See Fisher (1935), Dwass (1957), Good (2000), and Dufour (2006).

[63] Estimators have to be asymptotically linear, and the use of a studentized statistic, as defined in (6.113), is crucial for the asymptotic validity of the permutation approach (it would not be valid with the statistic $\tau = \hat{W}_x - \hat{W}_y$).

[64] Singh–Maddala distributions with parameters (a, q) equal to (2.5, 2.640350), (2.6, 2.218091), (2.7, 1.920967), (2.8, 1.7), (3.0, 1.3921126), (3.2, 1.1866026), (3.4, 1.0388049), (3.8, 0.8387663), (4.8, 0.5784599), and (5.8, 0.4473111) share the same (scale-invariant) Gini index, equal to 0.2887138. The tail parameters are, respectively, equal to $\alpha = 6.6$, 5.77, 5.19, 4.76, 4.18, 3.80, 3.53, 3.19, 2.78, and 2.59.

[65] For a two-tailed test, an asymptotic p-value is computed as $p_{as} = 2\min(\Phi(\tau); 1 - \Phi(\tau))$. A bootstrap p-value is similar, but bootstrap distribution replaces the asymptotic one, $p_{\text{boot}} = 2\min\left(\frac{1}{B}\sum_{b=1}^{B} \iota\left(\tau_b^* \leq \tau\right);\ \frac{1}{B}\sum_{b=1}^{B} \iota\left(\tau_b^* > \tau\right)\right)$. A permutation p-value is similar to the bootstrap p-value, p_{boot}, with B and τ_b^* replaced by P and τ_p^*. The null hypothesis is rejected if the p-value is smaller than the nominal level.

Table 6.8 Rejection frequencies of asymptotic, bootstrap, and permutation tests for testing the equality of Gini indices between two samples, when the underlying Singh–Maddala distributions are identical or different, $n=50$

	$\alpha_x=\alpha_y(F_x=F_y)$			$\alpha_x=4.76(F_x\neq F_y)$		
α_y	asym	boot	permut	asym	boot	permut
6.60	0.0704	0.0609	0.0515	0.0752	0.0624	0.0510
5.77	0.0727	0.0609	0.0504	0.0762	0.0634	0.0515
5.19	0.0751	0.0606	0.0512	0.0761	0.0622	0.0508
4.76	0.0770	0.0614	0.0500	0.0770	0.0614	0.0500
4.18	0.0828	0.0636	0.0499	0.0796	0.0612	0.0498
3.80	0.0871	0.0656	0.0497	0.0825	0.0639	0.0503
3.53	0.0938	0.0664	0.0494	0.0865	0.0668	0.0518
3.19	0.1015	0.0686	0.0508	0.0956	0.0719	0.0565
2.78	0.1154	0.0685	0.0515	0.1138	0.0824	0.0707
2.59	0.1277	0.0693	0.0513	0.1289	0.0911	0.0800

Table 6.8 shows the empirical rejection frequencies of asymptotic, bootstrap, and permutation tests for testing the equality of Gini indices between two samples, when the underlying distributions are identical and when they are different. As expected, when the distributions are identical, $F_x=F_y$, permutation tests provide exact inference, even in very small samples, with very heavy-tailed distributions (column 4: permut). When the distributions are different, $F_x\neq F_y$, permutation tests provide almost exact inference, except when F_y is much more heavy tailed than F_x (column 7: permut, when $\alpha_x=4.76$ and $\alpha_y\leq 2.78$). Overall, permutation tests outperform asymptotic and bootstrap tests.

Hill plots can be useful for studying the tail behavior in empirical studies. The "tail index" of a heavy-tailed distribution can be estimated by the Hill estimator on the k greatest order statistics of a sample of size n, for $k\leq n$—see Equation (6.109). However, estimation results can be sensitive to the choice of k. Hill plots show the Hill estimate of the tail index as a function of the number k of the greatest order statistics used to compute it. An estimate of the tail index can be selected when the plot becomes stable about a horizontal straight line. For instance, Figure 6.15 shows Hill plots obtained from two samples of 1000 observations drawn from the Singh–Maddala distributions SM(y; 2.8, 0.193, 1.7) and SM(y; 5.8, 0.193, 0.447), with tail parameters, respectively, equal to 4.76 and 2.59, over the range 0.5% to 25% of the largest order statistic used to compute it, with 95% confidence intervals (in gray). It is clear from this figure that the second sample (right plot) comes from a much more heavy-tailed distribution than the first sample (left plot).[66]

[66] The Hill plot is not always revealing. The Hill estimator is designed for the Pareto distribution. But the Hill plot can be very volatile—and thus difficult to interpret—when the upper tail of the underlying distribution is far from a Pareto, see Resnick (1997, 2007).

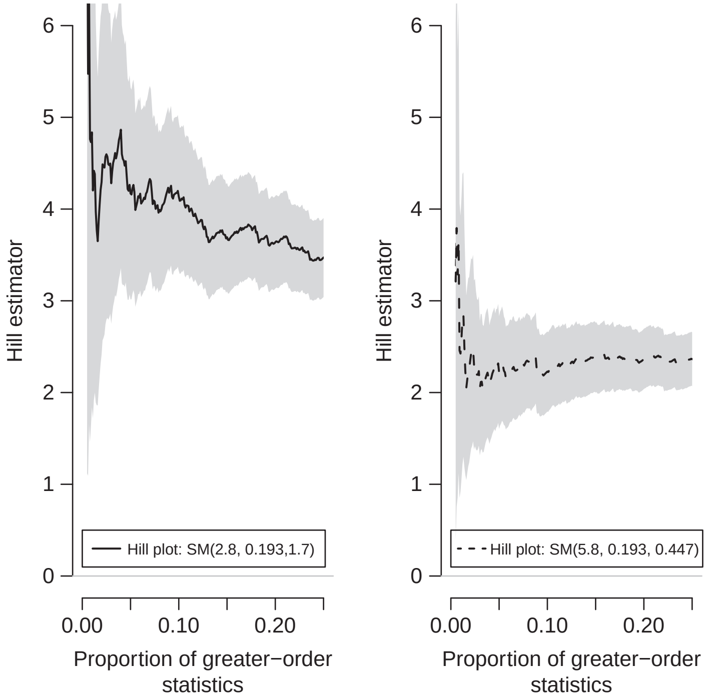


Figure 6.15 Hill plots: plot of the Hill estimate of the tail index (6.109) of two samples of 1000 observations drawn from Singh–Maddala distributions, as a function of the proportion of the greatest order statistics used to compute it.

6.4.6 Parametric Approaches

Sections 6.4.2–6.4.5 deal solely with distribution-free methods; in a sense we are working directly with the sample data. An alternative approach assumes that the distribution is known,[67] up to some parameters, and can be consistently estimated. A preliminary parametric estimation of the distribution is then obtained, and the moments of the parametric distribution are estimated. When the distribution is parametric, inequality indices can be expressed as functions of the distribution parameters. Table 6.9 shows the formulas of the Theil, MLD, and GE measures of inequality for the lognormal and Pareto distributions. They are also given for the Generalized Beta distribution of the second kind (GB2), a four-parameter distribution defined in Equation (6.5) when $c=1$, $\Gamma(\cdot)$ is the gamma function, and $\psi(\cdot):=\Gamma'(\cdot)/\Gamma(\cdot)$ is the digamma function.[68]

The Singh–Maddala distribution is the special case of the GB2 distribution when $p=1$; the Dagum distribution is the special case when $q=1$ (see Figure 6.6). Then, to

[67] See Section 6.3.1 for a discussion of the common functional forms that may be applied.
[68] See Jenkins (2009).

derive the expressions of the Theil, MLD, and GE indices for the Singh–Maddala and Dagum distributions, set $p=1$ and $q=1$ in the equations given in the last column in Table 6.9. An inequality measure can be estimated by replacing the unknown parameters by consistent parameter estimates. Inequality measures are then expressed as nonlinear functions of one or several consistent estimates. From the CLT, they are asymptotically Normal, and their asymptotic variance can be derived using the delta method.[69]

For several of the standard parametric distributions, the Gini index can also be expressed fairly easily as a function of the unknown parameters of the underlying distribution. The Gini index for the lognormal distribution, $\Lambda(y;\mu,\sigma^2)$, and for the Pareto distribution, $\Pi(y;\alpha)$, is equal to

$$I_{\text{Gini}}\left(\Lambda\left(y;\mu,\sigma^2\right)\right)=2\Phi\left(\frac{\sigma}{\sqrt{2}}\right)-1 \tag{6.117}$$

and

$$I_{\text{Gini}}=(\Pi(y;\alpha))=\frac{1}{2\alpha-1}, \tag{6.118}$$

respectively. For the Singh–Maddala distribution and for the Dagum distribution, defined in Equation (6.5) when $c=1$ and when, respectively, $p=1$ and $q=1$, the Gini index is equal to

Table 6.9 Parametric generalized entropy inequality measure for Lognormal (Λ), Pareto (Π), and generalized beta of the second kind (GB2) distributions

	$\Lambda(y;\mu,\sigma^2)$	$\Pi(y;\alpha)$	GB2$(y;a,b,p,q)$
I_{GE}^1	$\frac{\sigma^2}{2}$	$\frac{1}{\alpha-1}-\log\left(\frac{\alpha}{\alpha-1}\right)$	$-\log\left(\frac{\Gamma\left(p+\frac{1}{a}\right)\Gamma\left(q-\frac{1}{a}\right)}{\Gamma(p)\Gamma(q)}\right)+\frac{\psi\left(p+\frac{1}{a}\right)}{a}-\frac{\psi\left(q-\frac{1}{a}\right)}{a}$
I_{GE}^0	$\frac{\sigma^2}{2}$	$\log\left(\frac{\alpha}{\alpha-1}\right)-\frac{1}{\alpha}$	$\log\left(\frac{\Gamma\left(p+\frac{1}{a}\right)\Gamma\left(q-\frac{1}{a}\right)}{\Gamma(p)\Gamma(q)}\right)-\frac{\psi(p)}{a}+\frac{\psi(q)}{a}$
I_{GE}^ξ	$\frac{e^{(\xi^2-\xi)\frac{\sigma^2}{2}}-1}{\xi^2-\xi}$	$\frac{\frac{\alpha}{\alpha-\xi}\left(\frac{\alpha}{\alpha-1}\right)^{-\xi}-1}{\xi^2-\xi}$	$\frac{1}{\xi^2-\xi}\left(\frac{\Gamma\left(p+\frac{\xi}{a}\right)\Gamma\left(q-\frac{\xi}{a}\right)\Gamma^{\xi-1}(p)\Gamma^{\xi-1}(q)}{\Gamma^\xi\left(p+\frac{1}{a}\right)\Gamma^\xi\left(q-\frac{1}{a}\right)}-1\right)$

[69] Software-integrated commands can be used when calculations are cumbersome. Jenkins (2009) used the nlcom command in STATA to compute standard errors of GE indices for the GB2 distribution.

$$I_{\text{Gini}}(\text{SM}(y;a,b,q)) = 1 - \frac{\Gamma(q)\Gamma\left(2q - \frac{1}{a}\right)}{\Gamma\left(q - \frac{1}{a}\right)\Gamma(2q)} \tag{6.119}$$

and

$$I_{\text{Gini}}(\text{D}(y;a,b,p)) = \frac{\Gamma(p)\Gamma\left(2p + \frac{1}{a}\right)}{\Gamma(2p)\Gamma\left(p + \frac{1}{a}\right)} - 1 \tag{6.120}$$

The Singh–Maddala and Dagum distributions are encompassed by the Generalized Beta distribution of the second kind (GB2), defined in Equation (6.5) when $c=1$, for which the formula of the Gini index can also be obtained. However, its expression is lengthy and involves the generalized hypergeometric function, see McDonald (1984) or Kleiber and Kotz (2003) for an explicit formula. Because the Gini index is defined as nonlinear functions of one or several consistent estimates. From the CLT, it is asymptotically Normal, and the asymptotic variance can be derived using the delta method.

6.5. DISTRIBUTIONAL COMPARISONS

Apart from the simple welfare indices discussed in Section 6.4, we also need to be able to implement *ranking tools*. These tools provide the researcher with intuitively appealing methods of making distributional comparisons and are associated with important results in the welfare economics of distributional analysis.

6.5.1 Ranking and Dominance: Principles

The quantile and cumulation functionals Q and C defined in Section 6.4.1 can be used to establish *dominance criteria* for income distribution comparisons in terms of welfare or inequality, and related concepts are available for comparisons in terms of poverty.

6.5.1.1 Dominance and Welfare Indices

6.5.1.1.1 First-Order Dominance

Using (6.23), for a given $F \in \mathbb{F}$, the graph $\{q, Q(F,q) : q \in \mathbb{Q}\}$ describes Pen's parade (Pen, 1974). This is the basis for first-order distributional dominance (or first-order ranking) results. The concept of dominance can be explained as follows: consider two distributions $F, G \in \mathbb{F}$. Then F is said to *first-order dominate* G if the following pair of conditions hold:

$$\left.\begin{array}{l} \forall q \in \mathbb{Q} : Q(F,q) \geq Q(G,q), \\ \exists q \in \mathbb{Q} : Q(F,q) > Q(G,q). \end{array}\right\} \tag{6.121}$$

To see the importance of this concept, suppose we consider the class of all indices expressible in the form $W_{\mathrm{AD}}(F)$[70] additive social welfare functionals giving the aggregate of $\phi(y)$ where $\phi(\cdot)$ is some twice differentiable evaluation function of income. In particular, take the important subclass where welfare respects the monotonicity principle—the evaluation of income is everywhere strictly increasing:

$$\mathbb{W}_1 := \left\{ W | W(F) = \int \phi(y) \mathrm{d}F(y), \phi'(y) > 0 \right\}.$$

Then the statement "$W(F) \geq W(G)$, for any $W \in \mathbb{W}_1$" is equivalent to the statement "F first-order dominates G." If the Parade graph of F lies somewhere above and nowhere below the Parade graph of G, then welfare in F must be higher than in G, for any social welfare function that respects monotonicity (Quirk and Saposnik, 1962).

6.5.1.1.2 Second-Order Dominance

The functional (6.25) can be used to characterize a number of standard concepts associated with second-order dominance.

- For a given $F \in \mathbb{F}$, the graph $\{q, C(F, q) : q \in \mathbb{Q}\}$ describes the *generalized Lorenz curve* (GLC). This is the basis for second-order distributional dominance results (Shorrocks, 1983). The definition of second-order dominance can be derived from (6.121) just by replacing the quantile functional Q by the cumulant functional C. We also focus on a narrower subclass of welfare functions:

  $$\mathbb{W}_2 := \left\{ W | W(F) = \int \phi(y) \mathrm{d}F(y), \phi'(y) > 0, \phi''(y) \leq 0 \right\}.$$

  The concavity restriction $\phi''(y) \leq 0$ implies that a transfer of income from a poorer to a richer individual can never increase social welfare; it is a weak form of the transfer principle (Dalton, 1920). Then the statement "$W(F) \geq W(G)$, for any $W \in \mathbb{W}_2$" is equivalent to the statement "F second-order dominates G." If the GLC of F lies somewhere above and nowhere below the GLC of G, then welfare in F must be higher than in G, for any social welfare function that respects monotonicity and the transfer principle (Hadar and Russell, 1969). However, in distributional analysis attention is focused not only on the basic principle of second-order dominance, as just described, but also on restricted versions of this relationship that incorporate equivalence relationships on the members of $\mathbb{F}$.
- Suppose we want the second-order comparisons to be *scale independent*. This requires that, for any $F \in \mathbb{F}$ and any $\lambda > 0$ the distribution of y and of y/λ are regarded as equivalent for the purposes of distributional comparison; this implies that, when comparing distributions, we may divide incomes by an arbitrary positive constant. A natural

[70] See Equation (6.28)

choice for this constant is the mean of the distribution. The scale normalization of the GLC by the mean (6.26) gives the (*relative*) *Lorenz* functional:[71]

$$L(F;q) := \frac{C(F;q)}{\mu(F)} \tag{6.122}$$

and the graph $\{q, L(F,q) : q \in \mathbb{Q}\}$ gives the *relative Lorenz curve* (RLC).

- As an alternative to scale independence, we might be interested in a form of origin independence for the distributional comparisons, which would require that for any $F \in \mathbb{F}$ and any $\delta \in \mathbb{R}$ the distribution of y and of $y+\delta$ are regarded as equivalent. Instead of the scale normalization used in defining the RLC, we impose a "translational" normalization to define the absolute Lorenz curve. This is the graph $\{q, A(F;q) : q \in \mathbb{Q}\}$, where

$$A(F;q) := C(F;q) - q\mu(F).$$

6.5.1.2 Stochastic Dominance

The first-order and second-order dominance previously defined can be encompassed in a unified method, stochastic dominance, and extended to higher-order dominance.[72] Let us define dominance curves as follows:

$$D_F^s(y) := \frac{1}{(s-1)!}\int_0^y (y-t)^{s-1}\,\mathrm{d}F(t). \tag{6.123}$$

Distribution F is said to dominate distribution G stochastically at order s if the following pair of conditions holds:

$$\left.\begin{array}{l}\forall y \in \mathbb{Y} : D_F^s(y) \le D_G^s(y), \\ \exists y \in \mathbb{Y} : D_F^s(y) < D_G^s(y).\end{array}\right\} \tag{6.124}$$

The case $s=1$ corresponds to first-order dominance based on Pen's parade, previously defined in (6.121). Indeed, first-order stochastic dominance of G by F implies that $F(y) \le G(y)$ for all y, and there exists y over some interval for which the inequality holds strictly. It is similar to (6.121) expressed in terms of the quantile functions rather than CDFs.

The case $s=2$ corresponds to second-order dominance based on the GLC, previously defined. Indeed, from (6.123) and (6.25), we have $D_F^2(y_q) - D_G^2(y_q) = C(G;q) - C(F;q)$. Then, the pair of conditions in (6.124) is similar to that in (6.121) where the quantile functional Q is replaced by the cumulant functional C.

[71] This is equivalent to the income share (6.27).

[72] Fishburn (1980), O'Brien (1984), Stark and Yitzhaki (1988), Thistle (1989), O'Brien and Scarsini (1991), Fishburn and Lavalle (1995), and Davidson (2008).

There is a clear relation between dominance and poverty. From (6.86) and (6.123), we can see that $D_F^s(\zeta_0)$ is equal to the FGT poverty index, up to a scale factor. If, for all $[\zeta_0^-;\zeta_0^+]$, $D_F^s(\zeta_0) < D_G^s(\zeta_0)$, it follows that the FGT poverty index is lower in F than in G for all poverty lines in the interval $[\zeta_0^-;\zeta_0^+]$. The poverty measure can then be viewed as restricted stochastic dominance of G by F over that interval. The stochastic dominance criterion can also be viewed as a generalization of poverty measures when we let the poverty line vary over the whole support of the distribution.[73]

6.5.2 Ranking and Dominance: Implementation

To implement ranking criteria empirically, a standard approach is as follows[74]:

1. Choose a finite collection of population proportions $\Theta \subset \mathbb{Q}$.
2. For each $q \in \Theta$ compute the sample quantiles $\hat{y}_q$ and income cumulations $\hat{c}_q$ required for empirical implementation of first- and second-order rankings. To do this, we replace F in (6.23) and (6.25) by the EDF $F^{(n)}$—see Equation (6.8). Then we have

$$\hat{y}_q := Q\left(F^{(n)}; q\right) = y_{(\kappa(n,\, q))}, \tag{6.125}$$

where

$$\kappa(n, q) := \lfloor nq - q + 1 \rfloor \tag{6.126}$$

and $\lfloor x \rfloor$ denotes the largest integer no greater than x; we also have

$$\hat{c}_q := C\left(F^{(n)}; q\right) = \frac{1}{n} \sum_{i=1}^{\kappa(n,\, q)} y_{(i)} \tag{6.127}$$

3. Compute the variances and covariances of the sample quantiles (first-order) or the income cumulations (second-order).
4. Specify carefully the ranking hypothesis that is to be tested.

Step 1—involves a choice of how many points to select on the Parade or on the Lorenz curve. Step 2 is easy. Step 3 is dealt with in Section 6.5.2.1 and Step 4 in Sections 6.5.2.3 and 6.5.2.4.

6.5.2.1 Asymptotic Distributions

The main results follow from applying Lemmas in Section 6.4.2.2. We also need to define one further functional analogous to (6.23) and (6.25):

$$S(F; q) := \int_{\underline{y}}^{y_q} y^2 \mathrm{d}F(y) =: s_q. \tag{6.128}$$

and its sample counterpart:

[73] See Atkinson (1987) and Foster and Shorrocks (1988).

[74] We consider distribution-free methods. For parametric Lorenz curve comparisons, see Sarabia (2008).

$$\hat{s}_q := S\left(F^{(n)}; q\right) = \frac{1}{n}\sum_{i=1}^{\kappa(n,q)} y_{(i)}^2 \tag{6.129}$$

Then we have the following two theorems:

Theorem 1

For any $q, q' \in \mathbb{Q}$, $\sqrt{n}\hat{y}_q$ and $\sqrt{n}\hat{y}_{q'}$ *are asymptotically normally distributed with covariance*[75]:

$$\frac{q[1-q']}{f(y_q)f(y_{q'})}. \tag{6.130}$$

Proof

Immediate from Lemmas 1 and 2 ▪

Theorem 2

For any $q, q' \in \mathbb{Q}$, $\sqrt{n}\hat{c}_q$ and $\sqrt{n}\hat{c}_{q'}$ *are asymptotically normally distributed with covariance*[76]:

$$\omega_{qq'} := s_q + [qy_q - c_q]\left[y_{q'} - q'y_{q'} + c_{q'}\right] - y_q c_q \ \text{ for } q \leq q'. \tag{6.131}$$

Proof

Using Lemmas 1 and 3, the asymptotic covariance *of* $\sqrt{n}C\left(F^{(n)}; q\right)$ and $\sqrt{n}C\left(F^{(n)}; q'\right)$ is given by

$$\begin{aligned} \omega_{qq'} &= \int \mathrm{IF}(z; C(F;q), F)\mathrm{IF}(z; C(F;q'), F)\mathrm{d}F(z) \\ &= \int [qy_q - c_q + \iota(y_q \geq z)[z - y_q]]\left[q'y_{q'} - c_{q'} + \iota(y_{q'} \geq z)[z - y_{q'}]\right]\mathrm{d}F(z) \end{aligned} \tag{6.132}$$

Given that $\iota(x_{q'} \geq z) = 1$ whenever $\iota(x_q \geq z) = 1$ the right-hand side becomes

$$\begin{aligned} &[qy_q - c_q]\left[q'y_{q'} - c_{q'}\right] + \int_{\underline{y}}^{y_{q'}} [qy_q - c_q]\left[z - y_{q'}\right]\mathrm{d}F(z) \\ &+ \int_{\underline{y}}^{y_q} \left[q'y_{q'} - c_{q'} + z - y_{q'}\right][z - y_q]\mathrm{d}F(z) \end{aligned} \tag{6.133}$$

Using the definitions in (6.23), (6.25), and (6.128), we find that (6.133) becomes (6.131) ▪

We can also rewrite the IF as a random variable minus its expectation. From (6.42) in Lemma 3, we have

$$\mathrm{IF}(z; C(F,q), F) = Z_q - E(Z_q) \ \text{ where } Z_q = [z - y_q]\iota(z \leq y_q). \tag{6.134}$$

[75] See Lemma 1 of Beach and Davidson (1983).

[76] See Theorem 1 of Beach and Davidson (1983).

From (6.132), we can see immediately that the asymptotic covariance of $\sqrt{n}\hat{c}_q$ and $\sqrt{n}\hat{c}_{q'}$ is equal to the covariance of Z_q and $Z_{q'}$

$$\omega_{qq'} = \text{cov}\left(Z_q, Z_{q'}\right). \tag{6.135}$$

From a sample y_i, for $i=1, \ldots, n$, we can then estimate the covariance of the GLC ordinates $\hat{c}_q$ and $\hat{c}_{q'}$ as the empirical covariance of Z_{iq} and $Z_{iq'}$ divided by n

$$\widehat{\text{cov}}\left(\hat{c}_q, \hat{c}_{q'}\right) = \frac{1}{n}\hat{\omega}_{qq'} = \frac{1}{n^2}\sum_{i=1}^{n}\left(Z_{iq} - \overline{Z}_q\right)\left(Z_{iq'} - \overline{Z}_{q'}\right) \tag{6.136}$$

where

$$Z_{iq} = \left[y_i - \hat{y}_q\right]\iota\left(y_i \leq \hat{y}_q\right) \tag{6.137}$$

and $\overline{Z}_q = n^{-1}\sum_{i=1}^{n} Z_{iq}$, $\hat{y}_q$ is given in (6.125).

In a practical implementation one replaces the individual components of the right-hand side of (6.131) by their sample counterparts[77] to obtain the following consistent estimate of $\omega_{qq'}$:

$$\hat{\omega}_{qq'} := \hat{s}_q + \left[q\hat{y}_q - \hat{c}_q\right]\left[\hat{y}_{q'} - q'\hat{y}_{q'} + \hat{c}_{q'}\right] - \hat{y}_q\hat{c}_q. \tag{6.138}$$

These results can also be used for the ordinates of the (relative) Lorenz curve. Using the standard result on limiting distributions of differentiable functions of random variables (Rao, 1973), or using the delta method in (6.60), the asymptotic covariances of $\sqrt{n}\hat{c}_q/\hat{\mu}$ and $\sqrt{n}\hat{c}_{q'}/\hat{\mu}$ are then given by

$$v_{qq'} = \frac{1}{\mu^4}\left[\mu^2\omega_{qq'} + c_q c_{q'}\omega_{11} - \mu c_q \omega_{g'1} - \mu c_{q'}\omega_{q1}\right] \quad \text{for } q \leq q'. \tag{6.139}$$

where $\omega_{q1} := s_q + [qy_q - c_q]\mu - y_q c_q$, $\omega_{11} := s_1 - \mu^2$ and $\mu = \mu(F)$. Again, for practical implementation, the components of the right-hand side of (6.139) are replaced by their sample counterparts. We can also use the IF and express it as a random variable minus its expectation. The IF of the Lorenz curve ordinate (6.122) is given by[78]

$$\text{IF}(z; L(F,q), F) = \frac{1}{\mu}\left[qy_q - \frac{zc_q}{\mu} + \left[z - y_q\right]\iota\left(z \leq y_q\right)\right] \tag{6.140}$$

We can rewrite the IF as $\text{IF}(z; L(F,q), F) = Z_q - E(Z_q)$, where

$$Z_q = \frac{1}{\mu^2}\left[\mu\left[z - y_q\right]\iota\left(z \leq y_q\right) - c_q z\right]. \tag{6.141}$$

[77] $\hat{y}_q$, $\hat{c}_q$, and $\hat{s}_q$ are given by (6.125), (6.127), and (6.129), respectively.

[78] See Cowell and Victoria-Feser (2002) and Donald et al. (2012).

The asymptotic covariance of $\sqrt{n}\hat{c}_q/\hat{\mu}$ and $\sqrt{n}\hat{c}_{q'}/\hat{\mu}$ is equal to the covariance of Z_q and $Z_{q'}$ From a sample y_i, for $i=1, \ldots, n$, *we* can then estimate the covariance of the (relative) Lorenz curve ordinates $\hat{c}_q/\hat{\mu}$ and $\hat{c}_{q'}/\hat{\mu}$ as the empirical covariance of Z_{iq} and $Z_{iq'}$ divided by n,

$$\widehat{\text{cov}}\left(\frac{\hat{c}_q}{\hat{\mu}}, \frac{\hat{c}_{q'}}{\hat{\mu}}\right) = \frac{1}{n}\hat{v}_{qq'} = \frac{1}{n^2}\sum_{i=1}^{n}\left(Z_{iq} - \overline{Z}_q\right)\left(Z_{iq'} - \overline{Z}_{q'}\right) \tag{6.142}$$

where

$$Z_{iq} = \frac{1}{\hat{\mu}^2}\left[\hat{\mu}\left[y_i - \hat{y}_q\right]\iota\left(y_i \le \hat{y}_q\right) - \hat{c}_q y_i\right]. \tag{6.143}$$

and $\overline{Z}_q = n^{-1}\sum_{i=1}^{n} Z_{iq}, \hat{y}_q$ and $\hat{c}_q$ are given by (6.125) and (6.127), respectively.

The case of stochastic dominance can also be considered. For a given value z, a consistent estimator of $D_F^s(z)$ is

$$\hat{D}_F^s(z) = \frac{1}{n(s-1)!}\sum_{i=1}^{n}(z - y_i)^{s-1}\iota(y_i \le z) \tag{6.144}$$

where y_i, $i=1, \ldots, n$ is a random sample of n independent observations. Because it is a sum of IID observations, this estimator is consistent and asymptotically normal. The asymptotic covariance is also easy to calculate.[79]

When we compare two distributions, random samples can be obtained from two independent populations or from two correlated populations. The last case typically occurs when the two samples are independent paired drawings from the same population, as, for instance, with pretax and posttax distributions. In both cases of independent and correlated samples, it can be shown that the difference $\hat{D}_F^s\left(z_q\right) - \hat{D}_G^s\left(z_{q'}\right)$ is asymptotically normal, with asymptotic covariance equal to

$$\frac{1}{((s-1)!)^2}E\left[\left(z_q - y_F\right)_+^{s-1}\left(z_{q'} - y_G\right)_+^{s-1}\right] - D_F^s\left(z_q\right)D_G^s\left(z_{q'}\right) \tag{6.145}$$

where $(x)_+^{s-1} = x^{s-1}\iota(x \ge 0)$. This result comes from the CLT, assuming that population moments of order $2s-2$ for each distribution exist. The asymptotic covariance can be estimated with sample counterparts, with the expectation in (6.145) replaced by

$$\frac{1}{n}\sum_{i=1}^{n}\left(z_q - y_{F^{(n)},i}\right)_+^{s-1}\left(z_{q'} - y_{G^{(n)},i}\right)_+^{s-1} \tag{6.146}$$

[79] It is equal to (6.145) with $F=G$.

and $D^s(x)$ estimated as defined in (6.144). For $s=2$, we find an estimate of the covariance matrix similar to that obtained in (6.136) and (6.137), for the GLC ordinates. More details and explicit expressions for z being stochastic and for poverty measures can be found in the comprehensive approach to inference on stochastic dominance presented in Davidson and Duclos (2000).[80]

6.5.2.2 Dominance: An Intuitive Application

Armed with Theorems 1 and 2, an intuitive approach to dominance can be immediately applied. Using (6.127) we can plot an empirical GLC with confidence bands. Consistent estimates of the variance of the GLC ordinates can be calculated using (6.136) and (6.137) with $q=q'$. Therefore, we can immediately construct an informative graphical presentation for distributional comparisons, $\left(q, \hat{c}_q\right)$, with 95% confidence bands computed as $\left[\hat{c}_q \pm 1.96 \times \widehat{\mathrm{var}}\left(\hat{c}_q\right)\right]$. One could see whether it is reasonable to conclude that the GLC for distribution F lies above that for distribution G (second-order dominance). Clearly the same idea could be pursued with empirical quantiles and parade diagrams (first-order dominance).

Figure 6.16a, to the left, shows two GLCs obtained from two independent samples of 5000 observations drawn from Singh–Maddala distributions F and G, respectively, with confidence bands at 95% evaluated at the percentiles, $q=0.01, 0.02, \ldots, 0.99$. F is the Singh–Maddala distribution with parameters $a=2.8$, $b=0.193$, and $q=1.7$, used in the introduction, and G is the distribution with parameters $a=3.8$, 0.193, and 0.839; the means are, respectively, 0.169 and 0.240. This figure shows that distribution G second-order dominates distribution F. It suggests that poverty measures based on poverty gaps will exhibit more poverty in F than in G (Jenkins and Lambert, 1997). Table 6.10 shows poverty measures computed from the two samples, with 95% confidence intervals (see Section 6.4.4). As expected, poverty indices are significantly greater in F than in G.

We can also plot an empirical (relative) Lorenz curve with confidence bands. Consistent estimates of the variance of the (relative) Lorenz curve ordinates, $\widehat{\mathrm{var}}\left(\hat{c}_q/\hat{\mu}\right)$, can be calculated using (6.142) and (6.143) with $q=q'$. Therefore, we can construct a graphical representation of Lorenz curves, $\left(q, \hat{c}_q/\hat{\mu}\right)$, with 95% confidence bands, $\left[\left(\hat{c}_q/\hat{\mu}\right) \pm 1.96 \times \widehat{\mathrm{var}}\left(\hat{c}_q/\hat{\mu}\right)\right]$. It is often difficult to distinguish between two RLCs by eye; for this reason a plot of the difference between two RLCs is often useful. When the samples are independent, the variance of the difference between the ordinates is the sum of the variances from each sample. In practice, (relative) Lorenz curves often

[80] On dominance with complex sample design, see Beach and Kaliski (1986) and Zheng (1999, 2002). For an alternative approach focusing on crossings in the tails of Lorenz curves, see Schluter and Trede (2002), and for a Bayesian approach, see Hasegawa and Kozumi (2003). On the extension to absolute dominance and deprivation dominance, see Bishop et al. (1988) and Xu and Osberg (1998), and on poverty dominance see also Chen and Duclos (2008) and Thuysbaert (2008).

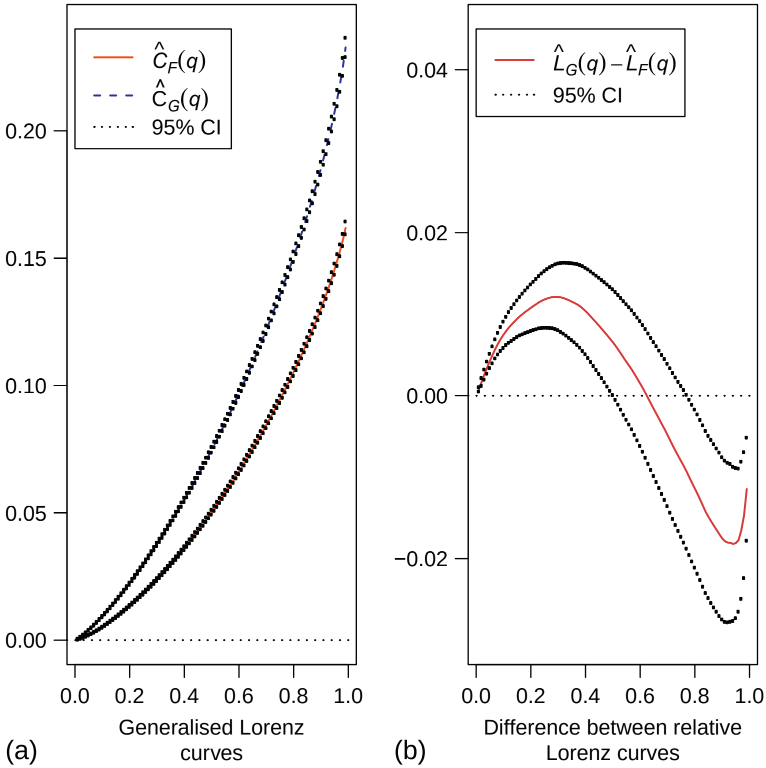


Figure 6.16 Generalized Lorenz curves and difference between Lorenz curves, $n = 5000$.

intersect, and in such cases no unambiguous ranking can be obtained. Nevertheless, useful information on inequality can be drawn from Lorenz curve comparisons.

Figure 6.16b, to the right, shows the difference between two RLCs obtained from the two samples used in Figure 6.16a. The two curves intersect in the upper part. This figure also shows that the empirical Lorenz curve of G lies significantly above the empirical Lorenz curve of F in the bottom part, whereas the reverse is true in the upper part. It suggests that inequality measures sensitive to the bottom part of income distributions would be smaller in G than in F, whereas inequality measures more sensitive to the upper part of income distributions would be smaller in F than in G. Table 6.10 shows inequality measures computed from the two samples, with 95% confidence intervals (see Section 6.4.3) Noting that GE measures, I^{ξ}_{GE}, are more sensitive to the bottom (top) of income distributions with smaller (higher) parameter ξ, we find the results suggested by the previous Lorenz curve comparisons. Indeed, I^{-1}_{GE} is significantly smaller in G than in F, whereas I^{2}_{GE} is significantly smaller in F than in G.

This approach is clearly *ad hoc*, and we need to examine the issues involved more carefully; we do this in Sections 6.5.2.3 and 6.5.2.4. Graphical representation of two

Table 6.10 Inequality and poverty measures, with confidence intervals at 95%, computed from two samples of 5000 observations drawn independently from *F* and *G*

	Distribution *F*		Distribution *G*	
	Index	**$CI_{95\%}$**	**Index**	**$CI_{95\%}$**
Poverty measures[a]				
P^0_{FGT}	0.1140	[0.1052;0.1228]	0.0180	[0.0143;0.0217]
P^1_{FGT}	0.0329	[0.0297;0.0360]	0.0038	[0.0028;0.0048]
P_{Sen}	0.0460	[0.0417;0.0503]	0.0055	[0.0041;0.0070]
P_{SST}	0.0635	[0.0575;0.0695]	0.0075	[0.0055;0.0096]
Generalized entropy measures				
I^{-1}_{GE}	0.1998	[0.1858;0.2137]	0.1551	[0.1455;0.1647]
I^0_{GE}	0.1520	[0.1448;0.1591]	0.1418	[0.1331;0.1506]
I^1_{GE}	0.1458	[0.1376;0.1539]	0.1564	[0.1435;0.1693]
I^2_{GE}	0.1693	[0.1538;0.1847]	0.2164	[0.1863;0.2465]
I_{Gini}	0.2937	[0.2869;0.3005]	0.2920	[0.2833;0.3007]

[a]The poverty line is half the median of the sample drawn from distribution *F*: $\zeta_0 = 0.07517397$.

empirical Lorenz curves, with confidence intervals, allows us to make *individual* comparisons. We can test if particular Lorenz curve ordinates are significantly different between two curves. To be able to make conclusions on dominance or nondominance, we need to test simultaneously that all ordinates from one curve are significantly greater or not smaller than the ordinates from the other curve. Appropriate test statistics need to be used to make *multiple* comparisons and to test simultaneously that several inequalities hold. Moreover, Lorenz curve ordinates are typically strongly positively correlated and, thus, test statistics need to take into account the covariance structure between the Lorenz curve ordinates.

6.5.2.3 The Null Hypothesis: Dominance or Nondominance

Performing inference on stochastic dominance is more complex than on a single welfare index. The hypotheses tested are usually based on a set of inequalities. For instance, first-order stochastic dominance requires that,

$$F(y) \leq G(y) \quad \text{for all } y \geq 0, \tag{6.147}$$

to say that distribution F dominates distribution G stochastically at order one. The theoretical literature also include the condition that $F(y) < G(y)$ for some y, as defined in (6.124). However, no statistical test can distinguish between these two forms of weak and strict dominance.[81] Because we are interested in statistical issues hereafter, we make

[81] Under the null of an inequality in one direction, a test cannot reject equality. Indeed, equality is on the frontier of the inequality hypothesis, and a test cannot distinguish statistically between being on the frontier and being very close to the frontier.

no distinction between weak and strict dominance, and we can write all inequalities as weak.

Inference on dominance in the population would be drawn from the corresponding sample properties. From a given sample, we can consistently estimate the two distributions by their EDF counterparts, $F^{(n)}(x)$ and $G^{(n)}(x)$. Sample dominance is then defined as $F^{(n)}(y) \leq G^{(n)}(y)$ for all y. It is clear that dominance in the population cannot be rejected if there is dominance in the sample. It is rejected if sample nondominance is statistically significant only. A similar reasoning applies for nondominance in the population. It follows that, to infer dominance, we should test the null hypothesis of nondominance, and to infer nondominance, we should test the null of dominance.

It can be illustrated with a simple example of two distributions with the same support and three points, $y_1 < y_2 < y_3$.[82] Because $F(y_3) = G(y_3) = 1$, we will say that distribution F dominates distribution G in the population if $d_i = G(y_i) - F(y_i) \geq 0$ for $i = 1, 2$. Figure 6.17 shows two bidimensional plots of $\hat{d}_1$ and $\hat{d}_2$, where the null hypothesis is,

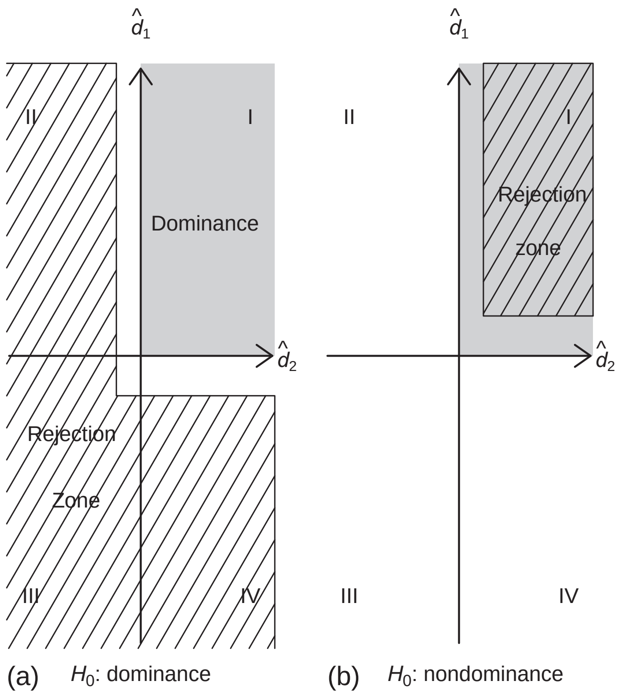


Figure 6.17 Tests of dominance and nondominance. The first quadrant, I, corresponds to dominance of G by F in the sample (gray area). The quadrants II, III, and IV correspond to nondominance.

[82] See Davidson and Duclos (2013).

respectively, dominance and nondominance. Distribution F dominates G in the sample when $\hat{d}_i \geq 0$, for $i=1, 2$. Then, the first quadrant, denoted I (gray area), corresponds to dominance in the sample, whereas quadrants II, III, and IV correspond to nondominance in the sample.

First, let us consider the null hypothesis of dominance, as shown in Figure 6.17a, on the left. To reject dominance in the population, the nondominance in the sample must be statistically significant, that is, the rejection zone has to be far enough from the dominance area, for example, in the cross-hatched area. The rejection zone is exclusively in the area of nondominance, whereas the (remaining) nonrejection zone corresponds to the dominance area plus the white L-shaped band within the nondominance area. Then, rejecting the null hypothesis of dominance corresponds to the case of nondominance, whereas nonrejecting it is inconclusive.

Second, let us consider the null hypothesis of nondominance, as shown in Figure 6.17b, on the right. With a similar reasoning, we can see that the rejection zone is exclusively in the area of dominance, whereas the nonrejection zone is composed of both nondominance and dominance situations (gray L-shaped band). Then, rejecting the null hypothesis of nondominance corresponds to the case of dominance, whereas nonrejecting it is inconclusive.

The previous example illustrates that positing the null of nondominance is the only way to draw a strong conclusion of dominance. However, it comes at a cost: Dominance will be inferred only if there is strong evidence in its favor. From Figure 6.17b, we can see that rejecting the null of nondominance is quite demanding because it requires that both statistics $\hat{d}_1$ and $\hat{d}_2$ are statistically significant. It may be too demanding, especially in the tails where both distributions tend to the same values and where we usually have sparse data and little information. Davidson and Duclos (2013) showed that, with distributions continuous in the tails, it is impossible to reject the null of nondominance over the full support of the distributions. It leads them to develop restricted stochastic dominance, limiting attention to some interval in the middle of the distribution.

The most common approach in the literature has developed tests of stochastic dominance positing the null of dominance.[83] The previous example illustrates the standard feature in statistics that nonrejecting the null does not imply that the null is true, and so selecting the null hypothesis remains allows for the possibility of being wrong at some level.[84] The level at which we may be wrong by accepting the null is unknown

[83] See Beach and Richmond (1985), McFadden (1989), Bishop et al. (1992), Anderson (1996), Schmid and Trede (1996), Davidson and Duclos (2000), Barrett and Donald (2003), Linton et al. (2005), and Maasoumi and Heshmati (2008).

[84] It is usually done when a coefficient is not significant in estimation results and when the analysis that follows is based on the selected regression model without the associated covariate.

(the L-shaped bands in Figure 6.17a and b), but it would be reduced by using statistical tests with greater power properties in a finite sample.

Finally, the two approaches can be seen as complementary. Rejecting the null of dominance or nondominance allows us to infer, respectively, nondominance and dominance when comparing two distributions.

6.5.2.4 Hypothesis Testing

Test statistics have been developed in the literature under the null hypothesis of dominance and nondominance. We distinguish both cases, for which we can interpret them, respectively, as union–intersection and intersection–union tests.

Under the Null of Dominance

Statistical tests can be constructed to test the null hypothesis of dominance against the alternative of nondominance. Under the null hypothesis that F dominates G, we have

$$\begin{aligned} H_0 &: D_F^s(y) \le D_G^s(y), \quad \text{for all } y \in \mathbb{Y}, \\ H_1 &: D_F^s(y) > D_G^s(y), \quad \text{for some } y \in \mathbb{Y}^{\cdot} \end{aligned} \tag{6.148}$$

where $\mathbb{Y}$ denotes a given set contained in the union of the support of the two distributions. An appropriate test statistic could be interpreted as a *union–intersection* test because the null hypothesis is expressed as an intersection of individual hypotheses and the alternative as an union (Roy, 1953). A natural test is based on the supremum of individual differences,

$$\tau = \sup_{y \in \mathbb{Y}} \left(\hat{D}_F^s(y) - \hat{D}_G^s(y) \right). \tag{6.149}$$

It is clear that the null hypothesis is rejected if τ is significant and positive. McFadden (1989) proposed a test based on (6.149) for two independent samples of IID observations. For $s = 1$, it is a variant of the Kolmogorov–Smirnov statistic, with known properties. For $s = 2$, the asymptotic distribution under the null is not tractable. Barrett and Donald (2003) proposed simulation-based methods for estimating critical values, taking into account comparisons at all points of the support (functional approach) rather than at a fixed number of arbitrarily chosen points. Linton et al. (2005) proposed to use subsampling methods, permitting to estimate critical values in general settings, with arbitrary order s, dependent observations, and continuous and discrete supports. For multiple comparisons restricted to a fixed number of points $(y_1, \ldots, y_T)$, a Wald test of inequality restrictions can also be used. Let us note the covariance matrix estimates of $\hat{D}_F^s$ and $\hat{D}_G^s$, respectively, as $\hat{\Omega}_F$ and $\hat{\Omega}_G$, the Wald test statistic is computed by solving

$$\min_{\delta \ge 0} \left(\hat{D}_F^s - \hat{D}_G^s - \delta \right)^\top \left(\hat{\Omega}_F + \hat{\Omega}_G \right) \left(\hat{D}_F^s - \hat{D}_G^s - \delta \right). \tag{6.150}$$

The statistic is obtained by using an algorithm to solve quadratic programming problems. The distribution of the statistic is a mixture of chi-square with weights that require simulation methods to be consistently estimated (Dardanoni and Forcina, 1999).

Lorenz dominance can be tested using similar methods. Bishop et al. (1989) and Davidson and Duclos (1997) proposed a test for a fixed number of points,[85] whereas Donald and Barrett (2004) and Bhattacharya (2007) have considered versions of Lorenz dominance tests in a functional approach, taking into account comparisons at all points of the supports.

Under the Null of Nondominance

Other statistical tests have been developed to test the null hypothesis of nondominance against the alternative of dominance. Under the null that F does not dominate G, we have

$$\begin{aligned} H_0 &: D_F^s(y) \geq D_G^s(y), \quad \text{for some } y \in \mathbb{Y}, \\ H_1 &: D_F^s(y) < D_G^s(y), \quad \text{for all } y \in \mathbb{Y} \end{aligned}. \tag{6.151}$$

An appropriate test could be interpreted as an *intersection–union* test because the null hypothesis is expressed as an union and the alternative as an intersection of individual hypotheses (Gleser, 1973). The idea behind the intersection–union method is that the null is rejected only if each of the individual hypotheses can be rejected. A natural test is based on the infimum of individual differences,

$$\tau' = \inf_{y \in \mathbb{Y}^r} \left(\hat{D}_G^s(y) - \hat{D}_F^s(y) \right). \tag{6.152}$$

It is clear that the null hypothesis is rejected if τ' is significant and positive. The statistic τ' has to be defined over $\mathbb{Y}^r$, some closed interval contained in the interior of the joint support of the two distributions $\mathbb{Y}$. The main reason is that the null hypothesis would never be rejected if we consider the tails of the distributions, where data are sparse and where the differences between the two distributions tend to zero. Specifically, $\mathbb{Y}^r$ should be a restricted interval in $\mathbb{Y}$ that removes the tails of the distributions. Kaur et al. (1994) proposed a test based on (6.152) for $s=2$ with independent samples and continuous distributions F and G. Critical values can be taken from the Normal distribution, making the test easy to implement. However, it can have low power properties (Dardanoni and Forcina, 1999). Davidson and Duclos (2013) and Davidson (2009b) proposed a test for higher order, for correlated samples as well as uncorrelated samples, and for continuous and discrete distributions. They also showed that appropriate bootstrap methods permit researchers to obtain much better finite sample properties.

[85] See also Bishop et al. (1991a,b, 1992).

6.6. OTHER ESTIMATION PROBLEMS

In Sections 6.4 and 6.5, we assumed that data are always drawn from a representative sample of the whole population. For some researchers this state of affairs is something of a luxury. In this section, we discuss a number of common problems that need to be taken into account in practical application and the statistical methods of dealing with them.

6.6.1 Contamination

By data contamination we mean a set of observations that do not "belong" to the sample—see Section 6.2.3. The essentials of the formal approach can be explained using a simple model based on the distribution function (in fact, we have already seen the elements of this model in a different context—see Section 6.4.2). The idea is that, instead of observing a distribution $F \in \mathbb{F}$ directly, one sees it after it has been mixed with another distribution that represents contamination. The elementary model of this is presented in Equation (6.33), where one observes a distribution G given by

$$G = [1 - \delta]F + \delta H^{(z)} \tag{6.153}$$

where δ represents the proportion of contamination in the mixture that we observe and $H^{(z)}$ is the elementary "contamination distribution" (6.32), a single point mass at $z \in \mathbb{Y}$. From this minimalist structure, one can easily develop more interesting specifications of the model of contamination using a mixture of F with a distribution that is richer than $H^{(z)}$. A number of questions immediately arise: Does contamination matter in analyzing income distributions? How does contamination affect distributional comparisons? How may one appropriately estimate models of income distribution if there is reason to believe that contamination is an important issue?

6.6.1.1 The Concept of Robustness

To address the question "how important," we can use the tool introduced in the discussion of asymptotic inference (Section 6.4.2). The IF is precisely designed to gauge the sensitivity of a statistic to contamination. Consider some statistic T (for example, an inequality measure, a poverty index, or a Lorenz ordinate); then IF quantifies the impact of an infinitesimal amount of contamination on the statistic T, namely, $\frac{\partial}{\partial \delta} T(G)|_{\delta \to \infty}$ (assuming that T is differentiable)—see Hampel (1971, 1974) and Hampel et al. (1986). Clearly the size of this differential will depend on the exact specification of the contamination distribution: in the context of the elementary model (6.153) this would mean that it will depend on the exact location in $\mathbb{Y}$ of the point z (where the contamination is concentrated). Of particular interest are cases where this IF is unbounded for some value of z; the interpretation of this is that the statistic T is highly sensitive to an infinitesimal amount of contamination at point z. In the present context this is precisely what we mean by saying that a statistic is *nonrobust*; obviously if the IF for the statistic T is bounded for all values of z then it makes sense to describe T as a *robust* statistic. We will come to some examples of robust and nonrobust statistics in a moment. However, first it is worth making the commonsense point that even if we are only using robust statistics in our analysis, this does not mean that we can ignore the possibility of data contamination; in practice, it may be that the assumption that δ is vanishingly small is just unreasonable.

6.6.1.2 Robustness, Welfare Indices, and Distributional Comparisons

Does contamination matter for the tools that we discussed in Sections 6.4 and 6.5?

Basic cases. First, take two statistics whose properties can be easily deduced, the mean and median. Using the definition of the mixture distribution (6.153) with point-contamination (6.32) and the linearity of the mean functional, we can write the mean of the observed mixture distribution as

$$\mu(G) = \mu\left([1-\delta]F + \delta H^{(z)}\right) = [1-\delta]\mu(F) + \delta\mu\left(H^{(z)}\right) \tag{6.154}$$

Evaluating (6.154) for the elementary point-contamination distribution (6.32), we obtain

$$\mu(G) = [1-\delta]\mu(F) + \delta z. \tag{6.155}$$

The observed mean is a simple weighted sum (with weights $1-\delta$, δ) of the true mean $\mu(F)$ and the value of z where the contamination is concentrated. Now differentiate (6.155) with respect to δ, and we find the IF for the functional μ as follows:

$$\mathrm{IF}(z; \mu, F) = z - \mu(F). \tag{6.156}$$

It is easy to see from (6.156) that $\mathrm{IF}(z; \mu, F)$ is unbounded as z tends to $-\infty$ or $+\infty$: The mean is a nonrobust statistic. So if you want to use the mean as a welfare index, then the introduction of a very small amount of contamination sufficiently far out in one of the tails of the distribution will cause the observed value of the mean to be pulled away from the true value.[86] Now consider the median, as a particular case of the quantile functional (6.23); using the basic result Lemma 2 and setting $q = 0.5$ to obtain the median we have

$$\mathrm{IF}(z; Q(\cdot, 0.5), F) = \frac{q - \iota(q \geq F(z))}{f(Q(F, 0.5))} = \frac{q - \iota(y_{0.5} \geq z)}{f(y_{0.5})}. \tag{6.157}$$

[86] Using (6.42), the same type of reasoning can be used to show that Lorenz ordinates are also nonrobust (Cowell and Victoria-Feser, 2002).

It is clear that, as long as there is positive density at the median $y_{0\cdot 5}$, the IF in (6.157) is bounded (Cowell and Victoria-Feser, 2002). So, in contrast to the mean, the median is robust. The intuition is clear: if you throw a single alien observation into the formula for the mean then, if that observation is large enough, it can have a huge effect when averaged in with the other sample values. But the median simply marks the "halfway" point in the distribution: If you introduce a single alien observation to the right of the median, then the size of that observation (how far it is to the right of the median) has no effect on the observed halfway point.

Inequality. It turns out that most commonly used inequality indices behave in a way that is similar to the mean: they are nonrobust (Cowell and Victoria-Feser, 1996b). To see why, let us check the properties of the W_{QAD} class of welfare indices (6.30) on which many standard inequality measures are based. The IF for a typical member of this class is

$$\varphi(z,\mu(F)) - W_{\text{QAD}}(F) + [z-\mu(F)]\int \varphi_\mu(z,\mu(F))\mathrm{d}F(z) \tag{6.158}$$

where $\varphi(y,\mu(F))$ is the evaluation of each individual income y used in the formula (6.30). It is clear that contamination could have an impact through more than one route—there is the direct effect from the evaluation of z, the first term in (6.158); there is also an indirect route through the effect on the mean, the third term in (6.158). Notice that this indirect route contains the expression, $z-\mu(F)$, as the right-hand side of (6.156). From this we can see that if $\varphi_\mu(z,\mu(F))$ is not everywhere zero, contamination will cause quasi-additive welfare indices to be nonrobust. Now consider the direct route: Clearly if $\varphi(z,\mu(F))$ is unbounded as z approaches infinity or as z approaches zero, the particular index in the W_{QAD} class will be nonrobust; this is precisely what happens with nearly all commonly used inequality measures.[87] Why does this happen? Inequality measures are usually designed to be sensitive to extreme values at one or other end of the distribution, so placing a tiny amount of contamination sufficiently far out in one of the tails is going to have a big impact on measured inequality because of its built-in sensitivity. As an example, take the GE measures. From Equations (6.49)–(6.51) we see that

$$\varphi(z,\mu(F)) = \frac{[z/\mu(F)]^\xi - 1}{\xi^2-\xi}.$$

Clearly this is unbounded for $\xi \geq 0$ as $z\to\infty$ and is unbounded for $\xi \leq 0$ as $z\to 0$, so the inequality indices are nonrobust for contamination among very high incomes in the case

[87] The implication of this is that even with a richer model of contamination than the elementary (6.32) leaves the mean unchanged, quasi-additively decomposable inequality indices will be nonrobust (Cowell and Victoria-Feser, 1996b).

of top-sensitive members of the GE family and for contamination near zero in the case of bottom-sensitive members of the GE family.[88]

Poverty. By contrast, conventional poverty indices such as the FGT class (6.86) and the Sen index (6.93) are robust if the poverty line is exogenous or is a function of a robust statistic such as the median (Cowell and Victoria-Feser, 1996a). Again, the intuition is straightforward. From (6.79), the IF for an additively decomposable poverty measure with a fixed poverty line ζ_0 is

$$\mathrm{IF}(z; P, F) = p(z, \zeta_0) - P(F)$$

where $p(\cdot)$ is the poverty evaluation function. From (6.86), we can see that for the FGT class

$$p(z, \zeta_0) = [\max(1 - z/\zeta_0, 0)]^{\xi}$$

so that $p(0,\zeta_0) = 1, p(z,\zeta_0)$ is nonincreasing in z for $z < \zeta_0$ and $p(z,\zeta_0)$ for $z \geq \zeta_0$. In plain language contamination at the very bottom of the distribution (below the poverty line) has an impact that it is bounded below, but a very high observation has no effect on poverty, whether that observation is a genuine high income or contamination. Poverty measures such as the FGT class are robust under contamination.

6.6.1.3 Model Estimation

If inequality measures are typically nonrobust, what is to be done about the possibility of contamination? A potentially useful approach is to use a parametric functional form $f(y;\theta)$ to model all or part of the income distribution and then compute inequality from the modeled distribution. But of course the robustness property of the inequality index based on the modeled distribution will depend on the way the parameter vector $\theta \in \mathbb{R}^p$ is estimated. If one consider using maximum likelihood estimators (MLEs), for example, the robustness problem remains. Although MLEs are attractive in terms of their efficiency properties, they are usually nonrobust. If we consider the wider class of *M-estimators* characterized by[89]

$$\sum_{i=1}^{n} \psi(y_i; \theta) = 0 \tag{6.159}$$

where ψ is a function $\mathbb{R} \times \mathbb{R}^p \to \mathbb{R}^p$ one can find estimators with suitable robustness properties. These are the bounded-IF M-estimators with minimal asymptotic covariance matrix, known as *Optimal Bias-Robust Estimators* (OBRE)—see Huber (1981) and

[88] The logarithmic variance and the Gini coefficient are also nonrobust—see (6.71), and Cowell and Victoria-Feser (1996b), Cowell and Flachaire (2007).

[89] The MLE belong to the class (6.159); in this case ψ is equal to the score function.

Hampel et al. (1986). One can see OBRE as the solution to a trade-off between efficiency and robustness.

A standard way of specifying the OBRE is as follows. Fix a bound $c \geq \sqrt{p}$ on the IF; then the OBRE are defined as the solution in θ of

$$\sum_{i=1}^{n} \psi(x_i;\theta) = \sum_{i=1}^{n} [s(x_i;\theta) - a(\theta)] \cdot W_c(x_i;\theta) = 0 \tag{6.160}$$

where $s(x;\theta) = \frac{\partial}{\partial\theta} \log f(x;\theta)$, the score function, and

$$W_c(x;\theta) = \min \left\{ 1; \frac{c}{\| A(\theta)[s(x;\theta) - a(\theta)] \|} \right\} \tag{6.161}$$

and $W_c(x;\theta)$ is a weight imputed to each observation according to its influence on the estimator. The $p \times p$ matrix $A(\theta)$ and $a(\theta) \in \mathbb{R}^p$ are defined by

$$E\left[\psi(x;\theta)\psi(x;\theta)^{\mathrm{T}}\right] = \left[A(\theta)^{\mathrm{T}} A(\theta)\right]^{-1} \tag{6.162}$$

and

$$E[\psi(x;\theta) = 0 \tag{6.163}$$

The constant c acts as a regulator between efficiency (high values of c) and robustness (low values of c). The solution of (6.160) must usually be found iteratively.[90]

6.6.2 Incomplete Data

We now turn to the problems of estimation and inference in a situation where, in part of the sample, some information is unavailable. As we noted in Section 6.2.3, this situation is sometimes imposed by data providers, sometimes created by researchers who are attempting to deal with problems of data contamination as discussed in Section 6.6.1.

6.6.2.1 Censored and Truncated Data

Here, we are dealing with the cases summarized in the first row of Table 6.2 in Section 6.2.3 in which we take $\underline{z}$ and $\overline{z}$ as fixed boundaries.

Truncated data. For data represented by Case A in Table 6.2, inference can be approached as for inference in the complete-information case with a redefined population: The limits of the support of the distribution $(\underline{y}, \overline{y})$ are replaced by the narrower truncation limits $(\underline{z}, \overline{z})$. If we wish to say more, it may be possible to use a parametric method to estimate the truncated part of the distribution.

Censoring with minimal information. Now consider Case B in Table 6.2. Clearly if we do not use the observed point masses at $\underline{z}$ and $\overline{z}$, this could be just treated as Case A.

[90] See Victoria-Feser and Ronchetti (1994), Cowell and Victoria-Feser (1996b), and for grouped data, see Victoria-Feser and Ronchetti (1997).

However, if we want to do more, first-order comparisons can be carried out. We need the following statistics: $\overline{n}$ (the full sample size), $\overline{n}$ (the number of observations equal to $\underline{z}$), and $\overline{n}$ (the number of observations equal to $\underline{z}$).

Censoring with rich information. Clearly it is possible to do more in Case C than in the previous two cases: More welfare indices (for the whole population) can be handled. Depending on the richness of information in the censored parts, it may be possible to carry out inference on Lorenz curve ordinates and some welfare indices. First, if, in addition to the information described in the discussion of Case B, the means of the censored parts of the sample are given,[91] then second-order rankings and the Gini coefficient can be estimated. Then it makes sense to define the following:

$$\hat{c}_{\text{low}} := \frac{1}{n}\sum_{i=1}^{\underline{n}} y_{(i)},$$

$$\hat{c}_{\text{high}} := \frac{1}{n}\sum_{i=n-\overline{n}+1}^{n} y_{(i)}.$$

Inference may also be possible using the same methodology as for the complete data case. To do this, we would additionally need the following information

$$\hat{s}_{\text{low}} := \frac{1}{n}\sum_{i=1}^{\underline{n}} y^2{}_{(i)}$$

$$\hat{s}_{\text{high}} := \frac{1}{n}\sum_{i=n-\overline{n}+1}^{n} y^2{}_{(i)}$$

If these variance terms from the excluded portion of the sample are also made available, then the asymptotic variances and covariances for the income cumulations (GLC ordinates) for $q, q' \in (\underline{\beta}, \overline{\beta})$ are found as follows. Replace (6.127) and (6.129) by the following

$$\hat{c}_q := \hat{c}_{\text{low}} + \frac{1}{n}\sum_{i=\kappa(n,\underline{\beta})+1}^{\kappa(n,\,q)} y_{(i)} \tag{6.164}$$

$$\hat{s}_q := \hat{s}_{\text{low}} + \frac{1}{n}\sum_{i=\kappa(n,\underline{\beta})+1}^{\kappa(n,\,q)} y^2{}_{(i)} \tag{6.165}$$

and plug into (6.138). To compute asymptotic variance for the (relative) Lorenz curve and the Gini coefficient, we also need the following:

$$\hat{\mu} = \hat{c}_{\overline{\beta}} + \hat{c}_{\text{high}},\ \hat{s}_1 := \hat{s}_{\overline{\beta}} + \hat{s}_{\text{high}},\ \hat{\omega}_{q1} := \hat{s}_q + \left[q\hat{y}_q - \hat{c}_q\right]\hat{\mu} - y_q\hat{c}_q,\ \hat{\omega}_{11} := \hat{s}_1 - \hat{\mu}^2$$

[91] In some cases means will be available from data providers.

Comparing the outcome from these computations with the full-information case in Section 6.5.2.1, we can draw two important conclusions (Cowell and Victoria-Feser, 2003). First, if the necessary information about the censored part of the distribution is used, the standard errors are the same as in the full information case. Second, when the information about the censored part is not available, the standard errors are smaller.

6.6.2.2 Trimmed Data

In the case of trimmed data, a fixed proportion of the sample is discarded—see the second row of Table 6.2. The trimmed samples for computing welfare indices and making distributional comparisons are usually based on robustness arguments (Cowell and Victoria-Feser, 2006): outliers may seriously bias the point estimates as well as the variances of the distributional statistics that are of interest—see the discussion in Section 6.6.1.

Here, we assume that a given proportion $\underline{\beta}$ has been removed from the bottom of the distribution and $1-\overline{\beta}$ has been removed from the top. If $(y_{\underline{\beta}}, y_{\overline{\beta}})$ denotes the range of the trimmed sample values, then $y_{\underline{\beta}}$ and $y_{\overline{\beta}}$ are *random*. Because of this, and in contrast to the discussion of truncated and censored data (Section 6.6.2.1), case D in Table 6.2 requires more extensive reworking of the full-information analysis in Section 6.5.2.

Inference is carried out on the full distribution[92] conditional on the fact that known proportions have been trimmed from the tails. The *trimmed distribution* $\widetilde{F}_\beta$ is defined as:

$$\widetilde{F}_\beta(y) := \begin{cases} 0 & \text{if } y < Q(F\underline{\beta}) \\ b[F(y)-\beta] & \text{if } Q(F,\underline{\beta}) \le y < Q\left(F,\overline{\beta}\right). \\ 1 & \text{if } y \ge Q\left(F,\overline{\beta}\right) \end{cases} \tag{6.166}$$

where $b := 1/\left[\overline{\beta}-\underline{\beta}\right]$. Using (6.166) the β-trimmed counterparts to (6.25) and (6.128) the income cumulations are given by

$$c_{\beta,q} := C\left(\widetilde{F}_{\beta;q}\right) = b\int_{y_{\underline{\beta}}}^{y_q} y\,\mathrm{d}F(y), \tag{6.167}$$

$$s_{\beta,q} := S\left(\widetilde{F}_{\beta;q}\right) = b\int_{y_{\underline{\beta}}}^{y_q} y^2\,\mathrm{d}F(y) \tag{6.168}$$

[92] Given that the integration of IF. IF$^{\mathrm{T}}$ is required over the full distribution to derive the asymptotic covariance matrix, this might appear to invalidate the applicability of nonparametric techniques because of the lack of information on the structure of the trimmed data. Cowell and Victoria-Feser (2003) show that this supposition is groundless.

and the counterpart of (6.26) is given by $\mu_\beta := \mu(\widetilde{F}_\beta)$. Once again, the sample analogues of (6.166)–(6.168) are obtained by replacing F by the empirical distribution $F^{(n)}$. For example, $c_{\beta,q}$ is estimated by[93]

$$\hat{c}_{\beta,q} := C\left(\widetilde{F}_\beta^{(n)}; q\right) = \frac{b}{n} \sum_{i=1}^{\kappa(n,q)} y_{(i)} \iota\left(i > \kappa\left(n, \underline{\beta}\right) + 1\right), \tag{6.169}$$

where $\{y_{(i)},\ i = 1, \ldots, n\}$ is the ordered sample, and μ_β is estimated by the mean of the trimmed sample

$$\hat{\mu}_\beta := \mu\left(\widetilde{F}_\beta^{(n)}\right) = \frac{b}{n} \sum_{i=1}^{n} y_{(i)} \iota\left(\kappa\left(n, \underline{\beta}\right) + 1 < i < \kappa\left(n, \overline{\beta}\right)\right). \tag{6.170}$$

Lorenz criteria. To apply second-order dominance criteria, we need to know the properties of the income cumulation for the trimmed distribution $\widetilde{F}_\beta$ and its empirical counterpart $\widetilde{F}_\beta^{(n)}$. The income cumulations based on the ordinary and trimmed distributions are related as follows:

$$C(\widetilde{F}_\beta; q) = b\left[C(F; q) - C\left(F; \underline{\beta}\right)\right], \tag{6.171}$$

from which it is clear that plotting Lorenz curves, GLCs, and so on is straightforward.

The estimation of the asymptotic covariance between $\sqrt{n}\hat{c}_{\beta,q}$ and $\sqrt{n}\hat{c}_{\beta,q'}$ follows as before, from an application of the IF. We need to evaluate

$$\int \mathrm{IF}\left(z; C(\cdot; q), \widetilde{F}_\beta\right) \mathrm{IF}\left(z; C(\cdot; q'), \widetilde{F}_\beta\right) \mathrm{d}F(z)$$

and then we may compare the results with those in the complete-information case.[94] Using the definition of the IF then (6.171) implies that IF for the cumulative income functional with trimmed data is[95]

$$\begin{aligned}
\mathrm{IF}\left(z; C(\cdot; q), \widetilde{F}_\beta\right) &= -c_{\beta,q} + b\left[q y_q - \underline{\beta} y_{\underline{\beta}} + \iota\left(y_q \ge z\right)\left[z - y_q\right] - \iota\left(y_{\underline{\beta}} \ge z\right)\left[z - y_{\underline{\beta}}\right]\right] \\
&= -c_{\beta,q} + b\left[q y_q - \underline{\beta} y_{\underline{\beta}} - \iota\left(y_q \ge z\right) y_q + \iota\left(y_{\underline{\beta}} \ge z\right) y_{\underline{\beta}}\right] + b\left[\iota\left(y_q \ge z\right) - \iota\left(y_{\underline{\beta}} \ge z\right)\right] z
\end{aligned} \tag{6.172}$$

Taking the mean of $\mathrm{IF}\left(z; C(\cdot; q), \widetilde{F}_\beta\right) \mathrm{IF}\left(z; C(\cdot; q'), \widetilde{F}_\beta\right)$ for each $z = y_i$ it is clear that no value of $z = y_i < y_{\underline{\beta}}$ or $z = y_i > y_{\overline{\beta}}$ will contribute to the value of (6.172).

[93] Note that at $q = \overline{\beta}$ for $\underline{\beta} = 1 - \overline{\beta}$ one gets the traditional trimmed mean, which generalizes the median as a robust estimator of location.

[94] In Lemma 1 $F(z)$ is estimated by $F^{(n)}$ so that the integral reduces to the mean over the sample.

[95] As before, $c_{\beta,q}$, $c_{\beta,q'}$, y_q, $y_{q'}$ and y_β can be estimated by their sample counterparts.

Assume that the set of population proportions satisfies $\Theta \subset [\underline{\beta}, \overline{\beta}]$. Then Equation (6.172) yields the following result (Cowell and Victoria-Feser, 2003):

Theorem 3

Given an original untrimmed sample of size n and lower and upper trimming proportions $\underline{\beta}, 1-\overline{\beta} \in \mathbb{Q}$, *for any* $q, q' \in \Theta$ *such that* $q \leq q'$ *the asymptotic covariance of* $\sqrt{n}\hat{c}_{\beta,q}$ *and* $\sqrt{n}\hat{c}_{\beta,q'}$ *is given by*

$$\varpi_{qq'} = b^2 \left[\omega_{qq'} + w_{\underline{\beta}\underline{\beta}} + w_{\underline{\beta}} q - \omega_{\underline{\beta}} q' \right]$$

where $\omega_{qq'}$ *is defined in* (6.131).

If we take the set of proportions $\Theta = \{ q_i = \underline{\beta} + i/n : i = 1, \ldots, n/b \}$ then $\varpi_{qq'}$ can be estimated by

$$\begin{aligned} \hat{\varpi}_{q_i q_j} = & \left[q_i y_{(i)} - \underline{\beta} y_{(1)} - \sum_{k=1}^{i} \frac{y_{(k)}}{b n_\beta} \right] \\ & \times \left[\left[1 - q_j\right] y_{(j)} - \left[1 - \underline{\beta}\right] y_{(1)} + \sum_{k=1}^{j} \frac{y_{(k)}}{b n_\beta} \right] \\ & - \sum_{k=1}^{i} \frac{y_{(i)} y_{(k)} - y_{(k)}^2}{b n_\beta} + y_{(1)} \left[q_i y_{(i)} - \underline{\beta} y_{(i)} - \sum_{k=1}^{i} \frac{y_{(i)}}{b n_\beta} \right] \end{aligned} \tag{6.173}$$

In the case of the Lorenz curve ordinates, the asymptotic covariance of $\sqrt{n}\hat{c}_{\beta,q}/\hat{\mu}_\beta$ and $\sqrt{n}\hat{c}_{\beta,q'}/\hat{\mu}_\beta$ is given by

$$v_{qq',\beta} = \frac{b^2}{\mu_\beta^4} \left[\mu_\beta^2 \varpi_{qq'} + c_{\beta,q} c_{\beta,q'} \varpi_{\overline{\beta}\overline{\beta}} - \mu_\beta c_{\beta,q} \varpi_{q'\overline{\beta}} - \mu_\beta c_{\beta,q'} \varpi_{q\overline{\beta}} \right]. \tag{6.174}$$

Compare this with (6.139).

QAD Welfare indices. To evaluate inequality and poverty indices we can again follow the method of Section 6.4, but perform the computations on the trimmed distribution $\widetilde{F}_\beta$ defined in (6.166)—once again ignoring the information on the excluded part of the sample. This means that the trimmed version of (6.30) becomes

$$W_{\mathrm{QAD}}\left(\widetilde{F}_\beta\right) = b \int \varphi\left(x, \mu\left(\widetilde{F}_\beta\right)\right) \mathrm{d}F(x) \tag{6.175}$$

The sample analogs of $W_{\mathrm{QAD}}\left(\widetilde{F}_\beta\right)$ in (6.175) are then given by

$$\hat{w}_{\mathrm{QAD},\beta} := W_{\mathrm{QAD}}\left(\widetilde{F}_\beta^{(n)}\right) := \frac{b}{n} \sum_{i=1}^{n} \varphi\left(y_{(i)}, \hat{\mu}_\beta\right) \iota\left(\kappa\left(n, \underline{\beta}\right) + 1 < i < \kappa\left(n, \overline{\beta}\right)\right) \tag{6.176}$$

which is the counterpart of (6.43) but applied to the trimmed sample. Evaluating the IF we have[96]

$$\begin{aligned}\mathrm{IF}\left(z; W_{\mathrm{QAD}}, \widetilde{F}_\beta\right) = b\varphi\left(\max\left(y_{\underline{\beta}}, \min\left(z, y_{\overline{\beta}}\right)\right), \mu\left(\widetilde{F}_\beta\right)\right) - W_{\mathrm{QAD}}\left(\widetilde{F}_\beta\right) \\ + b\mathrm{IF}\left(z, C\left(\cdot; \overline{\beta}\right), \widetilde{F}_\beta\right)\int_{\mathrm{Q}\left(F, \underline{\beta}\right)}^{\mathrm{Q}\left(F, \overline{\beta}\right)} \varphi_\mu\left(x, \mu\left(\widetilde{F}_\beta\right)\right)\mathrm{d}F(x)\end{aligned} \tag{6.177}$$

Once again, an estimate of the asymptotic variance of $\sqrt{n}W_{\mathrm{QAD}}\left(\widetilde{F}_\beta^{(n)}\right)$ can be easily obtained by computing the mean of squares of $\mathrm{IF}\left(z; W_{\mathrm{QAD}}, \widetilde{F}_\beta\right), z = y_i, i = 1, \ldots, n.$[97] Define the following distribution (corresponding to case E in Table 6.2):

$$F_\beta^*(y) := \begin{cases} 0 & \text{if } y < \mathrm{Q}\left(F, \underline{\beta}\right) \\ F(y) & \text{if } \mathrm{Q}\left(F, \underline{\beta}\right) \leq y < \mathrm{Q}\left(F, \overline{\beta}\right) \\ 1 & \text{if } y \geq \mathrm{Q}\left(F, \overline{\beta}\right) \end{cases}. \tag{6.178}$$

We can then state the following (Cowell and Victoria-Feser, 2003):

Theorem 4

The asymptotic variance of $\sqrt{n}W_{\mathrm{QAD}}\left(\widetilde{F}_\beta^{(n)}\right)$ *for the trimmed distribution* $\widetilde{F}_\beta$ *is*

$$\begin{aligned} & b^2\mathrm{var}\left(\varphi\left(x, \mu\left(\widetilde{F}_\beta\right)\right); F_\beta^*\right) \\ & + 2b^3\mathrm{cov}\left(x, \varphi\left(x, \mu\left(\widetilde{F}_\beta\right)\right); F_\beta^*\right)\int_{\mathrm{Q}(F, \underline{\beta})}^{\mathrm{Q}\left(F, \overline{\beta}\right)} \varphi_\mu\left(x, \mu\left(\widetilde{F}_\beta\right)\right)\mathrm{d}F(x) \\ & + b^4\mathrm{var}\left(x; F_\beta^*\right)\left[\int_{\mathrm{Q}(F, \underline{\beta})}^{\mathrm{Q}\left(F, \overline{\beta}\right)} \varphi_\mu\left(x, \mu\left(\widetilde{F}_\beta\right)\right)\mathrm{d}F(x)\right]^2 \end{aligned} \tag{6.179}$$

Note that in (6.179) the variance and covariance terms for the linear functionals are defined on the distribution F_β^* as opposed to the trimmed distribution (6.166). All the components of (6.179) can be estimated from the trimmed sample.

[96] To see this evaluate the mixture distribution and apply (6.34) to get

$$\begin{aligned} & - W_{\mathrm{QAD}}\left(\widetilde{F}_\beta\right) + b\varphi\left(z, \mu\left(\widetilde{F}_\beta\right)\right)\iota\left(z \leq y_{\overline{\beta}}\right)\iota\left(z \geq y_{\leftarrow\beta}\right) - b\varphi\left(y_{\overline{\beta}}, \mu\left(\widetilde{F}_\beta\right)\right)\iota\left(z \leq y_{\overline{\beta}}\right) \\ & + b\varphi\left(y_{\underline{\beta}}, \mu\left(\widetilde{F}_\beta\right)\right)\iota\left(z \leq y_{\underline{\beta}}\right) + bIF\left(z, C\left(\cdot; \overline{\beta}\right), \widetilde{F}_\beta\right)\int_{\mathrm{Q}\left(F, \underline{\beta}\right)}^{\mathrm{Q}\left(F, \overline{\beta}\right)} \varphi_\mu\left(x, \mu\left(\widetilde{F}_\beta\right)\right)\mathrm{d}F(x) \\ & + b\overline{\beta}\varphi\left(y_{\overline{\beta}}, \mu\left(\widetilde{F}_\beta\right)\right) - b\underline{\beta}\varphi\left(y_{\underline{\beta}}, \mu\left(\widetilde{F}_\beta\right)\right) \end{aligned}$$

where the first two lines follow by analogy with (6.45). The third line is found by considering the way the mixture distribution affects the limits of integration in (6.175) using Lemma 2. Rearranging gives (6.177).

[97] Notice that the contribution of $z = y_i < y_{\underline{\beta}}$ or $z = y_i > y_{\overline{\beta}}$ to (6.177) is nil.

The Gini coefficient. With trimmed data, the Gini coefficient can be expressed as

$$I_{\text{Gini}}\left(\widetilde{F}_\beta\right) = 1 - 2\int_{\underline{\beta}}^{\overline{\beta}} \frac{C\left(\widetilde{F}_\beta, q\right)}{C\left(\widetilde{F}_\beta, \overline{\beta}\right)} \mathrm{d}q. \tag{6.180}$$

Using the same procedure as before, we first evaluate the IF for the Gini coefficient with trimmed data as:

$$\begin{aligned}
\mathrm{IF}\left(z; I_{\text{Gini}}, \widetilde{F}_\beta\right) = {} & \frac{2}{\mu_\beta}\int_{\underline{\beta}}^{\overline{\beta}} c_{\beta,q}\mathrm{d}q - \frac{2b}{\mu_\beta}\left[\int_{\underline{\beta}}^{\overline{\beta}} q y_q \mathrm{d}q + \int_{\underline{\beta}}^{\overline{\beta}} \iota\left(y_q \geq z\right)\left[z - y_q\right]\mathrm{d}q\right] \\
& + \frac{2}{\mu_\beta}\left[\iota\left(y_{\underline{\beta}} \geq z\right)\left[z - y_{\underline{\beta}}\right] + \underline{\beta} y_{\underline{\beta}}\right] \\
& + \frac{2}{\mu_\beta^2}\int_{\underline{\beta}}^{\overline{\beta}} c_{\beta,q}\mathrm{d}q\left(-\mu_\beta + b\left[\overline{\beta} y_{\overline{\beta}} - \underline{\beta} y_{\underline{\beta}} + \iota\left(y_{\overline{\beta}} \geq z\right)\left[z - y_{\overline{\beta}}\right] - \iota\left(y_{\underline{\beta}} \geq z\right)\left[z - y_{\underline{\beta}}\right]\right]\right)
\end{aligned}$$

Using this or the results of Theorem 3, we can obtain[98]

Theorem 5

The asymptotic variance of $\sqrt{n} I_{\text{Gini}}\left(\widetilde{F}_\beta^{(n)}\right)$ *is* $4b^2\vartheta_\beta/\mu_\beta^4$ *where*

$$\vartheta_\beta = \mu_\beta^2 \int_{\underline{\beta}}^{\overline{\beta}} \int_{\underline{\beta}}^{q} \varpi_{q'q}\mathrm{d}q'\mathrm{d}q | \mu_\beta^2 \int_{\underline{\beta}}^{\overline{\beta}} \int_{q}^{\overline{\beta}} \varpi_{qq'}\mathrm{d}q\mathrm{d}q + \varpi_{\overline{\beta}\overline{\beta}}\left[\int_{\underline{\beta}}^{\overline{\beta}} c_{\beta,q}\mathrm{d}q\right]^2 - 2\mu_\beta \int_{\underline{\beta}}^{\overline{\beta}} c_{\beta,q}\mathrm{d}q \int_{\underline{\beta}}^{\overline{\beta}} \varpi_{q\overline{\beta}}\mathrm{d}q \tag{6.181}$$

The estimates of ϑ_β are found by making use of (6.173), with $\hat{\mu}_\beta$ being the trimmed sample mean (6.170).

6.6.3 Semiparametric Methods

The problems that we address here may have arisen from situations where the researcher has concerns about data contamination and robustness (see Section 6.6.1) or where the data provider has truncated or censored the data (see Section 6.6.2).[99]

The type of problem to be analyzed can be simplified if we restrict attention to one leading case. If the support of the income distribution is bounded below, then the problems with contaminated data are going to occur only in the upper tail of the distribution (Cowell and Victoria-Feser, 2002). It may be reasonable to use a parametric model for the upper tail of the distribution (modeled on a proportion $\beta \in \mathbb{Q}$ of upper incomes) and to use the EDF directly for the rest of the distribution (the remaining proportion the $1 - \beta$ of lower incomes). There are four main issues:

[98] For proof of $\mathrm{IF}\left(z; I_{\text{Gini}}, \widetilde{F}_\beta\right)$ and Theorem 5 see Cowell and Victoria-Feser (2003).

[99] This section draws on Cowell and Victoria-Feser (2007).

- What parametric model should be used for the tail?
- How should the model be estimated?
- How should the proportion β be chosen?
- What are the implications for welfare indices and dominance criteria?

6.6.3.1 The Model

The parametric model most commonly used for the upper tail is the Pareto distribution (6.2)—see the discussion in Section 6.3.1.1. In principle the Pareto model has two parameters: We suppose here that the parameter y_0 is determined by the $1-\beta$ quantile $Q(F;\,1-\beta)$ defined in (6.23); the dispersion parameter α is of special of interest and is to be estimated from the data.[100]

The *semiparametric distribution* is then

$$\widetilde{F}(y)=\begin{cases} F(y) & y\leq Q(F;1-\beta) \\ 1-\beta\left(\dfrac{y}{Q(F;1-\beta)}\right)^{-\alpha} & y>Q(F;1-\beta)\end{cases}. \tag{6.182}$$

For $y>Q(F;\,1-\beta)$, the density $\widetilde{f}$ is

$$\widetilde{f}(y;\alpha)=\beta\alpha Q(F;1-\beta)^{\alpha}y^{-\alpha-1}.$$

In particular

$$\widetilde{f}\left(y_{1-\beta};\alpha\right)=\frac{\beta\alpha}{y_{1-\beta}}. \tag{6.183}$$

6.6.3.2 Model Estimation

To estimate the Pareto model for the upper tail of the distribution, one could of course use the MLE but the MLE for the Pareto model is known to be sensitive to data contamination (Victoria-Feser and Ronchetti, 1994). Alternatively one could use OBRE as discussed in Section 6.6.1.3, with $p=1$. Given a sample $\{y_i,\, i=1,\,\ldots,\,n\}$ and a bound $c\geq 1$ on the IF, the OBRE are defined implicitly by the solution $\hat{\alpha}\left(\widetilde{F}\right)$ in

$$\int_{Q(F;1-\beta)}^{\infty}\psi\left(y;\hat{\alpha}\left(\widetilde{F}\right),Q(F;1-\beta)\right)\mathrm{d}\widetilde{F}(y)=0.$$

When ψ is the score function $s(y;\alpha,Q(F;1-\beta))=\frac{1}{\alpha}-\log(y)+\log\left(Q(F;1-\beta)\right)$ we get the MLE. We get the OBRE when

[100] For the results that follow, α is assumed to be >2 for the variance to exist.

$$\psi(y;\alpha)=[s(y;\alpha)-a(\alpha)]W_c(y;\alpha)$$

with

$$W_c(y;\alpha)=\min\left\{1;\frac{c}{||A(\alpha)[s(y;\alpha)-a(\alpha)||}\right\}. \tag{6.184}$$

$A(\alpha)$ and vector $a(\alpha)$ are defined implicitly by

$$\begin{array}{c} E[\psi(y;a)\psi'(y;a)]=\left[A(\alpha)'A(\alpha)\right]^{-1} \\ E[\psi(y;a)]=0 \end{array}.$$

As explained in Section 6.6.1.3, the constant c parameterizes the efficiency-robustness trade-off. A common method for choosing c is to choose an efficiency level (relative to that of MLE) and derive the corresponding value for c: for the Pareto model, a value of $c=2$ leads to an OBRE achieving approximately 85% efficiency.

6.6.3.3 Choice of β

Clearly one could adopt a heuristic approach selecting by eye the amount β of the upper tail to be replaced.

Alternatively one could use the robust approach in Dupuis and Victoria-Feser (2006), who develop a robust prediction error criterion by viewing the Pareto model as a regression model. Rearranging (6.2) or (6.4), one can represent the linear relationship between the log of the y and the log of the inverse CDF:

$$\log\left(\frac{y}{y_0}\right)=-\frac{1}{\alpha}\log(1-F(y;\alpha)).$$

Given a sample of ordered data $y_{(i)}$, the Pareto regression plot of log $(y_{(i)})$ versus $-\log\left(\frac{n+1-i}{n+1}\right)$, $i=1, \ldots, n$ can be used to detect graphically the point above which the plot yields a straight line.

6.6.3.4 Inequality and Dominance

The effect on inequality of semiparametric modeling is easy to see. For example, if we wish to see how the GE indices are affected, one substitutes $\widetilde{F}$—defined in (6.182)—into (6.49)–(6.51) to obtain $I_{GE}^{\xi}\left(\widetilde{F}\right)$. For first-order and second-order dominance results, we need to look once more at the quantile and cumulative income functionals.

The quantile functional obtained using (6.182) is given by

$$Q\left(\widetilde{F},q\right)=\begin{cases} Q(F,q) & q\le 1-\beta \\ Q(F;1-\beta)\left(\frac{1-q}{\beta}\right)^{-1/\hat{\alpha}\left(\widetilde{F}\right)} & q>1-\beta \end{cases} \tag{6.185}$$

The cumulative income functional becomes

$$C\left(\widetilde{F};q\right)=\begin{cases}\displaystyle\int_{\underline{z}}^{Q(F,\,q)} y\mathrm{d}F(y) & q\leq 1-\beta\\ \displaystyle\int_{\underline{z}}^{Q(F,1-\beta)} y\mathrm{d}F(y) & \\ \displaystyle +\beta\frac{\hat{\alpha}\left(\widetilde{F}\right)}{1-\hat{\alpha}\left(\widetilde{F}\right)}Q(F;1-\beta)\left[\left(\frac{1-q}{\beta}\right)^{\frac{\hat{\alpha}\left(\widetilde{F}\right)-1}{\hat{\alpha}\left(\widetilde{F}\right)}}-1\right] & q>1-\beta\end{cases}. \tag{6.186}$$

The graph of (6.186) gives the semiparametric GLC. The mean of the semiparametric distribution is given by (6.186) with $q=1$, namely,

$$\mu\left(\widetilde{F}\right)=\int_{\underline{z}}^{Q(F,1-\beta)} y\mathrm{d}F(y)-\beta Q(F;1-\beta)\frac{\hat{\alpha}\left(\widetilde{F}\right)}{1-\hat{\alpha}\left(\widetilde{F}\right)}=c_{1-\beta}-\beta y_{1-\beta}\frac{\hat{\alpha}}{1-\hat{\alpha}}. \tag{6.187}$$

So, using (6.186) and (6.187), the semiparametric Lorenz curve is just the graph of

$$L\left(\widetilde{F};q\right)=\frac{C\left(\widetilde{F};q\right)}{\mu\left(\widetilde{F}\right)}. \tag{6.188}$$

Estimates of the GLC and the Lorenz curve for the semiparametric model can be found by replacing F with $F^{(n)}$ in (6.182) to obtain

$$\widetilde{F}_{\beta}(y)=\begin{cases}F^{(n)}(y) & y\leq Q(F;1-\beta)\\ 1-\beta\left(\frac{y}{Q(F;1-\beta)}\right)^{-\alpha} & y>Q(F;1-\beta)\end{cases} \tag{6.189}$$

and replacing $\widetilde{F}$ by b $\hat{F}_{\beta}$ in (6.186)–(6.188).

To illustrate, consider the problem of comparing wealth distributions across countries where we are concerned that the upper tail of the distribution may be suffering from some sort of contamination. Table 6.11 shows the results of estimating a Pareto tail for the net-worth distributions of the UK, Sweden, and Canada around the time of the millennium. It employs two different methods of estimating α (OLS and robust) and three different values for the modeled proportion β (top 10%, top 5%, and top 1%).[101] It is clear that the difference between the two methods of estimation in computing $\hat{\alpha}$ can easily be as large as the differences in $\hat{\alpha}$ between countries; in the case where $\beta=0.10$ compare the OLS and robust estimates for the UK with the OLS estimate for Sweden.

Panel (a) of Figure 6.18 shows the two regression methods in detail for the case of the top 10% in the UK using a Pareto plot. It is clear that there are some high net-worth

[101] The data are from the Luxembourg Wealth Study, a harmonized database that facilitates international comparisons—see http://www.lisdatacenter.org/our-data/lws-database/. See Cowell (2013) for more detail of this example.

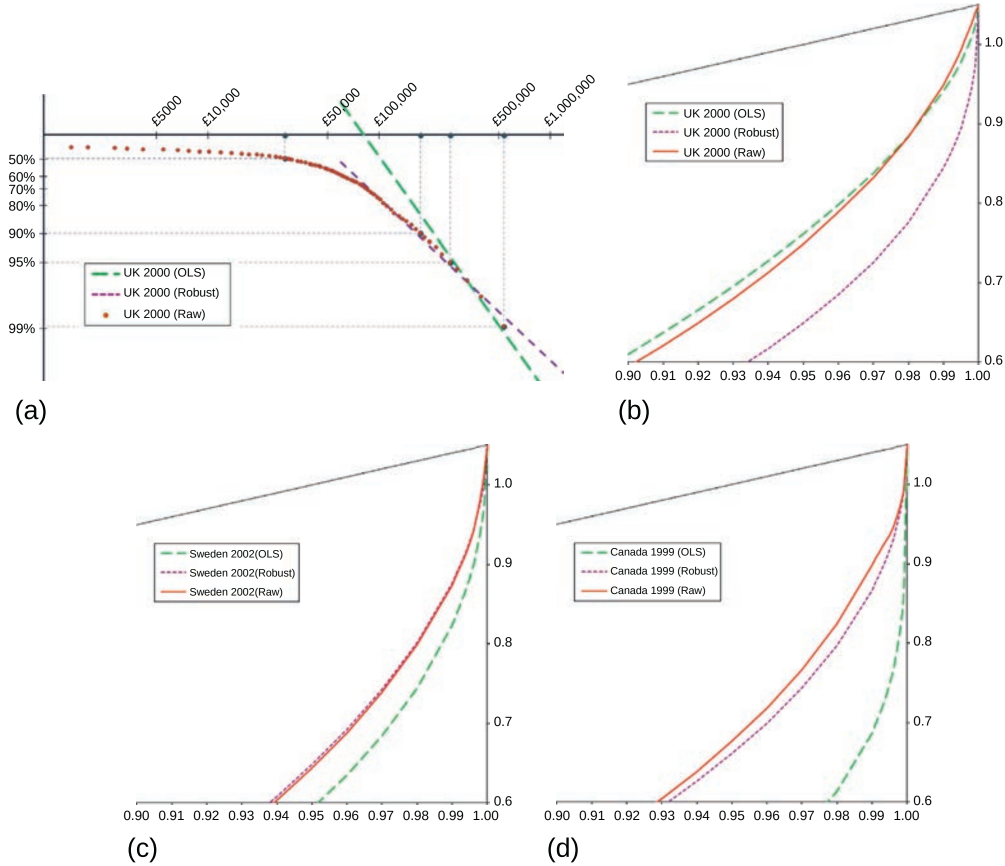


Figure 6.18 Semiparametric modeled Lorenz curves of net worth.

observations that "pull down" the OLS regression line so to speak; if one down-weights these observations, as in the robust regression, one finds a much flatter regression line, corresponding to a lower value of $\hat{\alpha}$ and, consequently, a higher estimate of inequality within the top 10% group. The results of OLS and robust methods used in semiparametric modeling are further illustrated for this case in Figure 6.18b, which shows the Lorenz curves for the raw data and for the semiparametric distributions produced by OLS and robust regression. Notice that if one considers the robust method appropriate, then the Lorenz curve for the whole distribution will lie well outside the Lorenz curve for the raw data and for the OLS semiparametric distribution.

Notice that the contrast between the OLS and robust estimates can differ dramatically between countries. This is evident from a comparison of the UK with Sweden or with Canada in Table 6.11. In the case of Sweden and Canada, the outliers pull the regression line in the opposite direction from that seen in the case of the UK: The robustly

Table 6.11 Estimates of α for net worth using different specifications of β

	$\beta = 0.10$	$\beta = 0.05$	$\beta = 0.01$
OLS estimation			
UK 2000	2.55	2.90	3.52
Sweden 2002	1.78	1.76	1.52
Canada 1999	1.37	1.53	1.94
Robust estimation			
UK 2000	1.71	2.08	3.07
Sweden 2002	2.10	2.18	1.61
Canada 1999	1.89	2.15	2.58

Source: Luxembourg Wealth Study.

Table 6.12 Estimates of the Gini coefficient for net worth from raw data and from semiparametric distributions

	$\hat{I}_{Gini}$	$I_{Gini}(\hat{F}_{0.10})$	$I_{Gini}(\hat{F}_{0.05})$	$I_{Gini}(\hat{F}_{0.01})$
OLS estimation				
UK 2000	0.665	0.657	0.660	0.665
Sweden 2002	0.893	0.901	0.901	0.902
Canada 1999	0.747	0.820	0.788	0.754
Robust estimation				
UK 2000	0.665	0.711	0.683	0.667
Sweden 2002	0.893	0.893	0.892	0.900
Canada 1999	0.747	0.752	0.747	0.745

Source: Luxembourg Wealth Study.

computed $\hat{\alpha}$ is accordingly higher than that found under OLS. The consequence for the Lorenz curves is shown in Figure 6.18c and d; it is clear that for Sweden and Canada the robustly estimated semiparametric Lorenz curve is close to the Lorenz curve for the raw data, but the OLS-estimated Lorenz curve is quite far away.

The effect of the different estimation methods on inequality *within* the top $100\beta\%$ is obvious—remember that the Gini coefficient for a Pareto distribution with parameter α is just $1/[2\alpha - 1]$. The resulting effect on Gini in the whole distribution is shown in Table 6.12. Although the effect can be quite large for $\beta = 0.10$, in none of the cases modeled here does it change the conclusion about the ranking by inequality of the three countries.

6.7. CONCLUSIONS

On reaching the end of a lengthy and technical chapter, the authors should confess to an uneasy feeling: A proportion of our potential readership might not have the stamina to work their way through every equation and every footnote. So, we would like to offer time-poor readers three things that may capture the essence of this chapter's contribution:

- A summary of lessons learned that we hope will be useful for practitioners and for other researchers;
- A little worked example that includes an application of many of the tools that we have discussed;
- A quick-reference table of the main formulas that should be useful to data providers as well as to the users of data.

6.7.1 Important Lessons: A Round-Up

Density Estimation, Parametric (Section 6.3.1)

(1) The Generalized Beta distribution encompasses all the standard parametric forms for income distribution. (2) A "good" goodness-of-fit criterion is important: Do use the Anderson–Darling statistic, the Cramér–von Mises statistic, or the Cowell et al. (2014) measure; do not use the χ^2 statistic

Density Estimation, Semi- and Nonparametric (Sections 6.3.2–6.3.4)

Standard kernel-density methods are very sensitive to the choice of the bandwidth. If the concentration of the data is markedly heterogeneous in the sample then the standard approach (the Silverman rule-of-thumb) is known to often oversmooth in parts of the distribution where the data are dense and undersmooth where the data are sparse, although in other cases it works well. However, this standard approach may not be suitable for income distributions, which are typically heavy-tailed: here the use of the adaptive kernel method or mixture model may be more appropriate.

Welfare Measures (Section 6.4)

(1) We propose a global approach to the derivation of variance expressions for all inequality measures. The method uses the IF (see Section 6.4.2.1) to provide a shortcut to the formulas we need. (2) It is necessary to analyze the tails (plot of Hill estimators) and use appropriate methods with heavy-tailed distributions (see Section 6.4.5.3).

Distributional Comparisons (Section 6.5)

(1) As with the welfare measures, we propose an approach to the variance and covariance formulas that again makes use of the IF. (2) A plot of Lorenz curve differences can provide useful information, even where Lorenz curves cross.

Data Problems (Section 6.6)

(1) Careful modeling is essential to understanding what can be done in the case of possible data contamination or incomplete data; again the IF is a valuable tool. (2) If one tries to "patch" an empirical distribution with a parametric model for the upper tail, then special attention needs to be given to the way the parameters of the model are to be estimated.

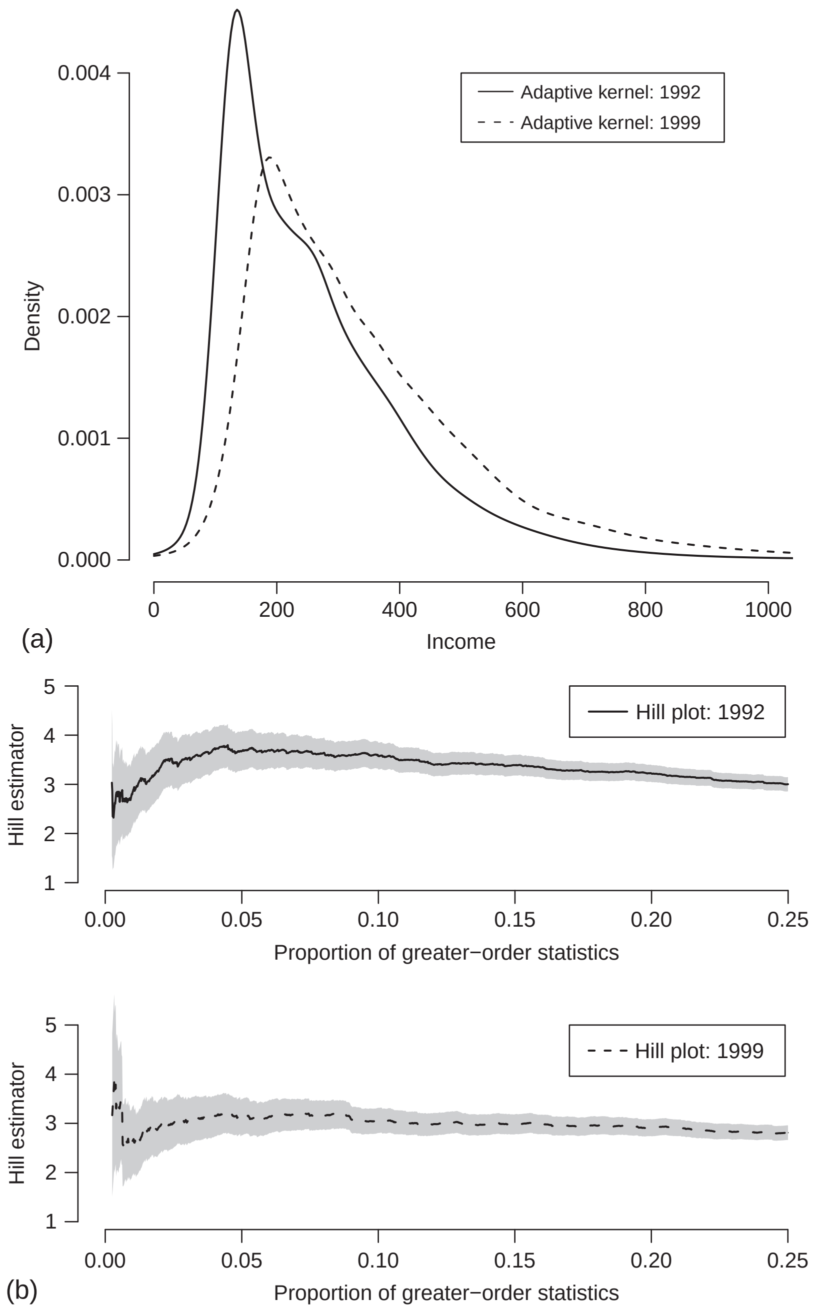
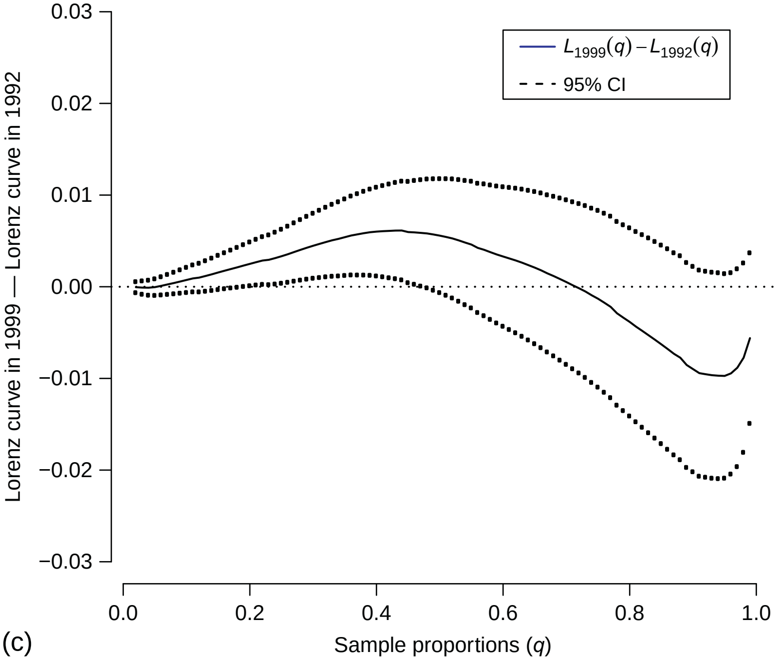


Figure 6.19 Inequality analysis on household income in 1992 and 1999 in United Kingdom: (a) Adaptive kernel density estimation, (b) Hill estimator of the tail index (Hill plots),

(Continued)



(c)

Index [a]	1992	1999	p-value [b]
I_{Gini}	0.3214	0.3200	0.7988
	[0.314;0.329]	[0.311;0.329]	
$I_{GE}^{-0.5}$	0.2279	0.2175	0.8288
	[0.190;0.369]	[0.186;0.378]	
I_{GE}^{0}	0.1742	0.1740	0.9529
	[0.166;0.184]	[0.163;0.185]	
I_{GE}^{1}	0.1794	0.1853	0.6006
	[0.167;0.194]	[0.170;0.203]	
I_{GE}^{2}	0.2415	0.2651	0.4244
	[0.210;0.288]	[0.227;0.336]	

[a] I_{GE}^{0}, I_{GE}^{1} and I_{GE}^{2} are, respectively, the MLD, the Theil index and half the square of the coefficient of variation.

[b] Permutation p-value for testing the equality of coefficients.

(d)

Figure 6.19—Cont'd (c) Difference of Lorenz curves, and (d) Inequality measures, with bootstrap confidence intervals and permutation p-values for testing equality.

6.7.2 A Worked Example

To illustrate these lessons, let us consider an empirical analysis of inequality measurement on the income distribution in the United Kingdom in 1992 and 1999.[102]

1. As noted in Section 6.7.1, income distributions are usually very skewed and heavy tailed, so fixed-bandwidth kernel density estimation, selected by Silverman's rule-of-thumb, may not be ideal (see Section 6.3.2). Figure 6.19a shows the application of one of the recommended methods, adaptive kernel density estimation (where the bandwidth varies with the degree of concentration of the data) of income distributions in 1992 and 1999.[103] The distribution in 1999 has a smaller mode and is shifted to the right, compared to 1992.
2. Statistical inference on inequality measures may be unreliable, in particular when the underlying distribution is quite heavy-tailed (see Sections 6.4.5.3 and 6.4.5.4). A Hill plot is a useful tool for studying the tail behavior in empirical studies: It represents the Hill estimator of the tail parameter, against the number of k-greatest order statistics used to compute it. An estimate of the tail parameter can be selected when the plot becomes stable about a horizontal straight line. Figure 6.19b shows Hill plots of income distribution in 1992 and 1999, over the range of 0.25% and 25% of order statistics used to compute it, with 95% confidence intervals (in gray). In 1992, the Hill estimate appears to be slightly more than 3, whereas it is very close to 3 in 1999. It suggests that the distribution in 1999 is slightly more heavy tailed than those in 1992, both being quite heavy tailed.[104]
3. Strong results on inequality ranking can be drawn from the comparison of RLCs, if the curves do not intersect (see Section 6.5). However, in empirical studies intersecting RLCs are not unusual, and we find that this is the case in our example, with the difference between the two Lorenz curves plotted in Figure 6.19c. The Lorenz curve for 1999 is above that for 1992 at the bottom of the distribution; the situation is reversed at the top of the distribution. It suggests that inequality measures more sensitive to transfers in the top (bottom) of the distribution would be larger (smaller) in 1999 than in 1992. However, the 95% confidence intervals shows that, at each point, Lorenz curve differences are not clearly statistically significant, and, thus inequality measures may not be statistically different in 1992 and 1999.

[102] The data are from the family expenditure survey, a continuous survey of samples of the UK population living in households. We take disposable household income before housing costs, divide household income by an adult-equivalence scale defined by McClements, and exclude the self-employed. The number of observations in 1992 and 1999 are, respectively, equal to 6597 and 5491.

[103] We obtain a very similar figure with an estimation based on a mixture of lognormal distributions. A kernel density estimation with a fixed bandwidth gives a slightly different picture, the difference being quite similar to that obtained in Figure 6.9.

[104] Note that the variance of a Pareto distribution exists if the Pareto index is >2.

Table 6.13 Formulas for computing coefficient estimates and variances for inequality measures, poverty measures, and (general or relative) Lorenz curve ordinates

Coefficient	Variance: $\widehat{\mathrm{var}}(\mathrm{Coef}) = \frac{1}{n}\mathrm{var}(Z) = \frac{1}{n^2}\sum_{i=1}^{n}(Z_i - \overline{Z})^2$ where Z_i is equal to
Inequality measures	
$\hat{I}_{\mathrm{GE}}^{\xi} = \frac{1}{n(\xi^2-\xi)}\sum_{i=1}^{n}\left[\left(\frac{y_i}{\hat{\mu}}\right)^{\xi} - 1\right]$	$Z_i = \frac{1}{\xi^2-\xi}\left(\frac{y_i}{\hat{\mu}}\right)^{\xi} - \xi\frac{y_i}{\hat{\mu}}\left[\hat{I}_{\mathrm{GE}}^{\xi} + \frac{1}{\xi^2-\xi}\right]$
$\hat{I}_{\mathrm{GE}}^{0} = -\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{y_i}{\hat{\mu}}\right)$	$Z_i = \frac{y_i}{\hat{\mu}} - \log y_i$
$\hat{I}_{\mathrm{GE}}^{1} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i}{\hat{\mu}}\right)\log\left(\frac{y_i}{\hat{\mu}}\right)$	$Z_i = \frac{y_i}{\hat{\mu}}\left[\log\left(\frac{y_i}{\hat{\mu}}\right) - \hat{I}_{\mathrm{GE}}^{1} - 1\right]$
$\hat{I}_{\mathrm{MD}} = \frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \hat{\mu}\rvert$	$Z_i = 2(\hat{q}-1)y_i + 2(y_i - \hat{\mu})\iota(y_i \geq \hat{\mu})$
$\hat{I}_{\mathrm{Gini}} = \frac{2\sum_{i=1}^{n} i y_{(i)}}{\hat{\mu}n(n-1)} - \frac{n+1}{n-1}$	$Z_i = \frac{1}{\hat{\mu}}\left[-\left(\hat{I}_{\mathrm{Gini}} + 1\right)y_{(i)} + \frac{2i-1}{n}\,y_{(i)} - \frac{2}{n}\sum_{j=1}^{i} y_{(j)}\right]$
Poverty measures	
$\hat{P}_{\mathrm{FGT}}^{\xi} = \frac{1}{n}\sum_{i=1}^{n} Z_i = \overline{Z}$	$Z_i = \left\lvert 1 - \frac{y}{\zeta_0}\right\rvert^{\xi}\iota(y \leq \zeta_0)$
$\hat{P}_{\mathrm{Sen}} = \frac{2}{nn_p\zeta_0}\sum_{i=1}^{n_p}\left(\zeta_0 - y_{(i)}\right)\left(n_p - i + \frac{1}{2}\right)$	$Z_i = \frac{2n}{\zeta_0 n_p}\left[\frac{\zeta_0}{2}\left(\frac{2n_p}{n} - \hat{P}_{\mathrm{Sen}}\right) - \frac{2n_p - 2i + 1}{2n}\,y_{(i)} - \frac{1}{n}\sum_{j=1}^{i} y_{(j)}\right]\iota\left(y_{(i)} \leq \zeta_0\right)$
$\hat{P}_{\mathrm{SST}} = \frac{2}{\zeta_0 n(n-1)}\sum_{i=1}^{n_p}\left(\zeta_0 - y_{(i)}\right)(n-i)$	$Z_i = \frac{2n}{\zeta_0(n-1)}\left[\zeta_0\left(1 - \frac{n_p}{n}\right) - \frac{2n-2i+1}{2n}\,y_{(i)} + \frac{1}{n}\sum_{j=1}^{n_p} y_{(j)} - \frac{1}{n}\sum_{j=1}^{i} y_{(j)}\right]\iota\left(y_{(i)} \leq \zeta_0\right)$
Lorenz curves	
$\hat{c}_q = \frac{1}{n}\sum_{i=1}^{\kappa(n,\,q)} y_{(i)}$	$Z_{iq} = \left[y_i - \hat{y}_q\right]\iota\left(y_i \leq \hat{y}_q\right)$
$\hat{c}_q/\hat{\mu} = \frac{1}{\hat{\mu}n}\sum_{i=1}^{\kappa(n,\,q)} y_{(i)}$	$Z_{iq} = \frac{1}{\hat{\mu}^2}\left[\hat{\mu}\left(y_i - \hat{y}_q\right)\iota\left(y_i \leq \hat{y}_q\right) - \hat{c}_q y_i\right]$

$Z = \{Z_1, \ldots, Z_n\}$ and $\overline{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i; \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y_i$ is the sample mean; $y_{(i)}$ is the ith order statistic of the sample; $\hat{q} = \frac{1}{n}\sum_{i=1}^{n}\iota(y_i \leq \hat{\mu}); \zeta_0$ is the poverty line; $n_p = \sum_{i=1}^{n}\iota(y_i \leq \zeta_0)$ is the number of poor; q is a sample proportion; $\kappa[n,\ q] = \lfloor nq - q + 1\rfloor$ is the largest integer no greater than $nq - q + 1$, and $\hat{y}_q = y_{(\kappa(n,\,q))}$ is a sample quantile. $\hat{I}_{\mathrm{GE}}^{0}$ and $\hat{I}_{\mathrm{GE}}^{1}$ are, respectively, the mean logarithmic deviation and the Theil inequality indices.

4. Several inequality measures are computed in Figure 6.19d: The Gini index and the GE measures with a sensitivity parameter equals to -0.5, 0, 1, 2. GE inequality measures are known to be more sensitive to transfers in the top (bottom) of the distribution as its parameter increases (decreases). Moreover, GE indices with parameters 0, 1, and 2 are, respectively, the MLD, the Theil and half the square of the coefficient of variation indices. Standard bootstrap confidence intervals are given in brackets. The two distributions are quite heavy tailed, but the tail parameters are not very different. Reliable inference for testing equality of coefficients can then be obtained with permutation tests (see Section 6.4.5.4): the p-values are given in the last column. The results show that the values of inequality measures that are more sensitive to the top (bottom) of the distribution are larger (smaller) in 1999 than in 1992. However, taking into account statistical inference leads us not to reject the hypothesis that the inequality measures are similar in 1992 and in 1999. These results are consistent with the previous analysis drawn from the Lorenz curves comparison.

6.7.3 A Cribsheet

Finally, we offer something for those who are really short of time or patience. In this chapter, we have proposed a unified approach for computing variances and covariances for many inequality and poverty measures, as well as Lorenz curve ordinates. This unified approach involves some quite simple—or at least not very complicated—formulas. Table 6.13 provides a one-page summary of the key formulas for the principal statistical tasks in distributional analysis.

ACKNOWLEDGMENTS

We thank Russell Davidson, Stephen Jenkins, Michel Lubrano, André-Marie Taptue, and the editors, Tony Atkinson and Francois Bourguignon, for helpful comments. Flachaire would like to thank McGill University and CIRANO for their hospitality and the Institut Universitaire de France and the Aix-Marseille School of Economics for financial support.

REFERENCES

Ahamada, I., Flachaire, E., 2010. Non-Parametric Econometrics. Oxford University Press, Oxford.

Aitchison, J., Brown, J.A.C., 1957. The Lognormal Distribution. Cambridge University Press, London.

Alvaredo, F., Saez, E., 2009. Income and wealth concentration in Spain from a historical and fiscal perspective. J. Eur. Econ. Assoc. 7 (5), 1140–1167.

Anderson, G., 1996. Nonparametric tests of stochastic dominance in income distributions. Econometrica 64, 1183–1193.

Atkinson, A.B., 1970. On the measurement of inequality. J. Econ. Theory 2, 244–263.

Atkinson, A.B., 1987. On the measurement of poverty. Econometrica 55, 749–764.

Autor, D.H., Katz, L.F., Kearney, M., 2008. Trends in U.S. wage inquality: revising the revisionists. Rev. Econ. Stat. 90 (2), 300–323.

Bandourian, R., McDonald, J.B., Turley, R.S., 2003. A comparison of parametric models of income distribution across countries and over time. Estadística 55, 135–152.
Bantilan, M.C.S., Bernal, N.B., de Castro, M.M., Pattugalan, J.M., 1995. Income distribution in the Philippines, 1957-1988: an application of the Dagum model to the family income and expenditure survey (FIES) data. In: Dagum, C., Lemmi, A. (Eds.), Income Distribution, Social Welfare, Inequality and Poverty. In: Research on Economic Inequality 6, JAI Press, Greenwich, pp. 11–43.
Barrett, G.F., Donald, S.G., 2003. Consistent tests for stochastic dominance. Econometrica 71, 71–104.
Barrett, G.F., Donald, S.G., 2009. Statistical inference with generalized Gini indices of inequality, poverty, and welfare. J. Bus. Econ. Stat. 27, 1–17.
Beach, C.M., Davidson, R., 1983. Distribution-free statistical inference with Lorenz curves and income shares. Rev. Econ. Stud. 50, 723–735.
Beach, C.M., Kaliski, S.F., 1986. Lorenz curve inference with sample weights: an application to the distribution of unemployment experience. Appl. Stat. 35 (1), 38–45.
Beach, C.M., Richmond, J., 1985. Joint confidence intervals for income shares and Lorenz curves. Int. Econ. Rev. 26 (6), 439–450.
Beran, R., 1988. Prepivoting test statistics: a bootstrap view of asymptotic refinements. J. Am. Stat. Assoc. 83, 687–697.
Berger, Y.G., Skinner, C.J., 2003. Variance estimation for a low income proportion. J. R. Stat. Soc.: Ser. C: Appl. Stat. 52 (4), 457–468.
Berndt, E.R., 1990. The Practice of Econometrics: Classic and Contemporary. Reading, Mass: Addison Wesley.
Bhattacharya, D., 2007. Inference on inequality from household survey data. J. Econ. 137, 674–707.
Biewen, M., Jenkins, S.P., 2006. Variance estimation for generalized entropy and atkinson inequality indices: the complex survey data case. Oxf. Bull. Econ. Stat. 68, 371–383.
Binder, D.A., Kovacevic, M.S., 1995. Estimating some measures of income inequality from survey data: an application of the estimating equations approach. Survey Methodol. 21 (2), 137–145.
Bishop, J.A., Chakraborti, S., Thistle, P.D., 1988. Large sample tests for absolute Lorenz dominance. Econ. Lett. 26, 291–294.
Bishop, J.A., Formby, J.P., Thistle, P.D., 1989. Statistical inference, income distributions and social welfare. In: Slottje, D.J. (Ed.), Research on Economic Inequality I. Connecticut: JAI Press, pp. 49–82.
Bishop, J.A., Formby, J.P., Smith, W.J., 1991a. International comparisons of income inequality: tests for Lorenz dominance across nine countries. Economica 58, 461–477.
Bishop, J.A., Formby, J.P., Smith, W.J., 1991b. Lorenz dominance and welfare: changes in the U.S. distribution of income, 1967-1986. Rev. Econ. Stat. 73, 134–139.
Bishop, J.A., Formby, J.P., Thistle, P., 1992. Convergence of the south and non-south income distributions, 1969-1979. Am. Econ. Rev. 82, 262–272.
Bishop, J.A., Formby, J.B., Zheng, B., 1997. Statistical inference and the Sen index of poverty. Int. Econ. Rev. 38, 381–387.
Boos, D.D., Serfling, R.J., 1980. A note on differential and the CLT and LIL for statistical functions, with application to *m*-estimates. Ann. Stat. 8, 618–624.
Bordley, R.F., McDonald, J.B., Mantrala, A., 1997. Something new, something old: parametric models for the size distribution of income. J. Income Distrib. 6, 91–103.
Bourguignon, F., Morrisson, C., 2002. Inequality among world citizens: 1820-1992. Am. Econ. Rev. 92, 727–744.
Bowman, A., 1984. An alternative method of cross-validation for the smoothing of kernel density estimates. Biometrika 71, 353–360.
Brachmann, K., Stich, A., Trede, M., 1996. Evaluating parametric income distribution models. Allg. Stat. Arch. 80, 285–298.
Burkhauser, R.V., Crews Cutts, A.D., Daly, M.C., Jenkins, S.P., 1999. Testing the significance of income distribution changes over the 1980s businesscycle: a cross-national comparison. J. Appl. Econ. 14, 253–272.
Burkhauser, R.V., Feng, S., Jenkins, S.P., 2009. Using the p90/p10 index to measure U.S. inequality trends with Current Population Survey data: a view from inside the Census Bureau vaults. Rev. Income Wealth 55 (1), 166–185.

Burkhauser, R.V., Feng, S., Larrimore, J., 2010. Improving imputations of top incomes in the public-use current population survey by using both cell-means and variances. Econ. Lett. 108 (1), 69–72.
Butler, R.J., McDonald, J.B., 1989. Using incomplete moments to measure inequality. J. Econ. 42, 109–119.
Chen, W., Duclos, J., 2008. Testing for Poverty Dominance: An Application to Canada, Discussion Paper 3829, IZA.
Chernozhukov, V., Fernandez-Val, I., Melly, B., 2009. Inference on Counterfactual Distributions, Working Paper.
Chesher, A., Schluter, C., 2002. Measurement error and inequality measurement. Rev. Econ. Stud. 69, 357–378.
Chotikapanich, D., Griffiths, W., 2008. Estimating income distributions using a mixture of gamma densities. In: Chotikapanich, D. (Ed.), Modelling Income Distributions and Lorenz Curves. Springer, New York, pp. 285–302, Chapter 16.
Chow, G.C., 1960. Tests of equality between sets of coefficients in two linear regressions. Econometrica 28, 591–605.
Chung, E., Romano, J.P., 2013. Exact and asymptotically robust permutation tests. Ann. Stat. 41, 484–507.
Cowell, F.A., 1989. Sampling variance and decomposable inequality measures. J. Econ. 42, 27–41.
Cowell, F.A., 1991. Grouping bounds for inequality measures under alternative informational assumptions. J. Econ. 48, 1–14.
Cowell, F.A., 1999. Estimation of inequality indices. In: Silber, J. (Ed.), Handbook on Income Inequality Measurement. Dewenter/Kluwer, Norwell, MA, pp. 269–290.
Cowell, F.A., 2000. Measurement of inequality. In: Atkinson, A.B., Bourguignon, F. (Eds.), Handbook of Income Distribution. Elsevier Science B. V., New York, pp. 87–166, Chapter 2.
Cowell, F.A., 2011. Measuring Inequality, third ed. Oxford University Press, Oxford.
Cowell, F.A., 2013. UK wealth inequality in international context. In: Hills, J.R. (Ed.), Wealth in the UK. Oxford University Press, Oxford, pp. 35–62. Chapter 3.
Cowell, F.A., Fiorio, C., 2011. Inequality decompositions—a reconciliation. J. Econ. Inequal. 9 (99), 509–528.
Cowell, F.A., Flachaire, E., 2007. Income distribution and inequality measurement: the problem of extreme values. J. Econ. 141, 1044–1072.
Cowell, F.A., Jenkins, S.P., 2003. Estimating welfare indices: household sample design. Res. Econ. Inequal. 9, 147–172.
Cowell, F.A., Victoria-Feser, M.-P., 1996a. Poverty measurement with contaminated data: a robust approach. Eur. Econ. Rev. 40, 1761–1771.
Cowell, F.A., Victoria-Feser, M.-P., 1996b. Robustness properties of inequality measures. Econometrica 64, 77–101.
Cowell, F.A., Victoria-Feser, M.-P., 2002. Welfare rankings in the presence of contaminated data. Econometrica 70, 1221–1233.
Cowell, F.A., Victoria-Feser, M.-P., 2003. Distribution-free inference for welfare indices under complete and incomplete information. J. Econ. Inequal. 1, 191–219.
Cowell, F.A., Victoria-Feser, M.-P., 2006. Distributional dominance with trimmed data. J. Bus. Econ. Stat. 24, 291–300.
Cowell, F.A., Victoria-Feser, M.-P., 2007. Robust stochastic dominance: a semi-parametric approach. J. Econ. Inequal. 5, 21–37.
Cowell, F.A., Jenkins, S.P., Litchfied, J.A., 1996. The changing shape of the UK income distribution: Kernel density estimates. In: Hills, J.R. (Ed.), New Inequalities: The Changing Distribution of Income and Wealth in the United Kingdom. Cambridge University Press, Cambridge, pp. 49–75, Chapter 3.
Cowell, F.A., Davidson, R., Flachaire, E., 2014. Goodness of Fit: an Axiomatic approach. J. Bus. Econ. Stat. forthcoming.
Cowell, F.A., Flachaire, E., Bandyopadhyay, S., 2013. Reference distributions and inequality measurement. J. Econ. Inequal. 11, 421–437.
Dagsvik, J., Jia, Z., Vatne, B.H., Zhu, W., 2013. Is the Pareto-Lévy law a good representation of income distributions? Empir. Econ. 44, 719–737.

Dagum, C., 1977. A new model of personal income distribution: specification and estimation. Econ. Appl. 30, 413–436.
Dagum, C., 1980. The generation and distribution of income, the Lorenz curve and the Gini ratio. Econ. Appl. 33, 327–367.
Dagum, C., 1983. Income distribution models. In: Banks, D.L., Read, C.B., Kotz, S. (Eds.), Encyclopedia of Statistical Sciences, vol. 4. Chichester, Wiley, New York, pp. 27–34.
Dalton, H., 1920. Measurement of the inequality of incomes. Econ. J. 30, 348–361.
Daly, M.C., Crews, A.D., Burkhauser, R.V., 1997. A new look at the distributional effects of economic growth during the 1980s: a comparative study of the United States and Germany. Econ. Rev. 2, 18–31.
Dardanoni, V., Forcina, A., 1999. Inference for Lorenz curve orderings. Econ. J. 2, 49–75.
Davidson, R., 2008. Stochastic dominance. In: Durlauf, S.N., Blume, L.E. (Eds.), The New Palgrave Dictionary of Economics. second ed, vol. 7. Palgrave Macmillan, Basingstoke, pp. 921–925.
Davidson, R., 2009a. Reliable inference for the Gini index. J. Econ. 150 (1), 30–40.
Davidson, R., 2009b. Testing for restricted stochastic dominance: some further results. Rev. Econ. Anal. 1, 34–59.
Davidson, R., 2010. Innis lecture: inference on income distributions. Can. J. Econ. 43, 1122–1148.
Davidson, R., 2012. Statistical inference in the presence of heavy tails. Econom. J. 15, C31–C53.
Davidson, R., Duclos, J.-Y., 1997. Statistical inference for the measurement of the incidence of taxes andtransfers. Econometrica 65, 1453–1466.
Davidson, R., Duclos, J.-Y., 2000. Statistical inference for stochastic dominance and for the measurement of poverty and inequality. Econometrica 68, 1435–1464.
Davidson, R., Duclos, J.-Y., 2013. Testing for restricted stochastic dominance. Econ. Rev. 32, 84–125.
Davidson, R., Flachaire, E., 2007. Asymptotic and bootstrap inference for inequality and poverty measures. J. Econ. 141, 141–166.
Davidson, R., MacKinnon, J.G., 2000. Bootstrap tests: how many bootstraps? Econ. Rev. 19, 55–68.
Davison, A.C., Hinkley, D.V., 1997. Bootstrap Methods. Cambridge University Press, Cambridge.
Deaton, A.S., 1997. The Analysis of Household Surveys. Johns Hopkins Press for the World Bank, Baltimore, Maryland.
Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via EM algorithm (with discussion). J. R. Stat. Soc. B 39, 1–38.
Deville, J.-C., 1999. Variance estimation for complex statistics and estimators: linearization and residual techniques. Surv. Methodol. 25, 193–203.
DiNardo, J., Fortin, N.M., Lemieux, T., 1996. Labor market institutions and the distribution of wages, 1973-1992: a semiparametric approach. Econometrica 64, 1001–1044.
Donald, S. G. and G. F. Barrett (2004). Consistent nonparametric tests for Lorenz dominance. Econometric Society 2004 Australasian Meetings 321, Econometric Society.
Donald, S.G., Green, D.A., Paarsch, H.J., 2000. Differences in wage distributions between Canada and the United States: an application of a flexible estimator of distribution functions in the presence of covariates. Rev. Econ. Stud. 67, 609–633.
Donald, S.G., Hsu, Y.-C., Barrett, G.F., 2012. Incorporating covariates in the measurement of welfare and inequality: methods and applications. Econom. J. 15, C1–C30.
Dufour, J.-M., 2006. Monte Carlo tests with nuisance parameters: a general approach to finite-sample inference and nonstandard asymptotics in econometrics. J. Econ. 133, 443–477.
Dufour, J.-M., Flachaire, E., Khalaf, L., 2013. Permutation tests for comparing inequality measures with heavy-tailed distributions. Technical report. In: Paper Presented at the 53th Conference of the Société Canadienne de Science Économique in Québec.
Dupuis, D.J., Victoria-Feser, M.-P., 2006. A robust prediction error criterion for Pareto modeling of upper tails. Can. J. Stat. 34 (4), 639–658. To appear.
Dwass, M., 1957. Modified randomization tests for nonparametric hypotheses. Ann. Math. Stat. 28, 181–187.
Efron, B., 1982. The Jackknife, the Bootstrap and Other Resampling Plans, vol. 38. Society for Industrial and Applied Mathematics, Philadelphia.

Epanechnikov, V.A., 1969. Nonparametric estimation of a multidimensional probability density. Theory Probab. Appl. 14, 153–158.
Escobar, M.D., West, M., 1995. Bayesian density estimation and inference using mixtures. J. Am. Stat. Assoc. 90, 577–588.
Fan, J., Yim, T.H., 2004. A cross-validation method for estimating conditional densities. Biometrika 91, 819–834.
Ferguson, T.S., 1983. Bayesian density estimation by mixtures of normal distributions. In: Rizvi, M.H., Rustagi, J., Siegmund, D. (Eds.), Recent Advances in Statistics: Papers in Honor of Herman Chernoff on His Sixtieth Birthday. Academic Press, New York, pp. 287–302.
Fernholz, L.T., 1983. Von Mises Calculus for Statistical Functionals. Lecture Notes in Statistics,19 Springer, New York.
Fishburn, P.C., 1980. Stochastic dominance and moments of distributions. Math. Oper. Res. 5, 94–100.
Fishburn, P.C., Lavalle, I.H., 1995. Stochastic dominance on unidimensional grids. Math. Oper. Res. 20, 513–525.
Fisher, R., 1935. The Design of Experiments. Oliver & Boyd, London.
Flachaire, E., Nuñez, O., 2007. Estimation of income distribution and detection of subpopulations: an explanatory model. Comput. Stat. Data Anal. 51, 3368–3380.
Foster, J.E., Shorrocks, A.F., 1988. Poverty orderings. Econometrica 56 (1), 173–177.
Foster, J.E., Greer, J., Thorbecke, E., 1984. A class of decomposable poverty measures. Econometrica 52, 761–776.
Frühwirth-Schnatter, S., 2006. Finite Mixture and Markov Switching Models. Springer, New York.
Gastwirth, J.L., 1971. A general definition of the Lorenz curve. Econometrica 39, 1037–1039.
Gastwirth, J.L., 1972. The estimation of the Lorenz curve and Gini index. Rev. Econ. Stat. 54, 306–316.
Gastwirth, J.L., 1974. Large-sample theory of some measures of inequality. Econometrica 42, 191–196.
Gastwirth, J.L., 1975. The estimation of a family of measures of economic inequality. J. Econ. 3, 61–70.
Gastwirth, J.L., Nayak, T.K., Krieger, A.N., 1986. Large sample theory for the bounds on the Gini and related indices from grouped data. J. Bus. Econ. Stat. 4, 269–273.
Ghosal, S., van der Vaart, A.W., 2001. Entropies and rates of convergence for maximum likelihood and bayes estimation for mixtures of normal densities. Ann. Stat. 29 (5), 1233–1263.
Gibrat, R., 1931. Les Inegalités Economiques. Sirey, Paris.
Giles, D.E.A., 2004. A convenient method of computing the Gini index and its standard error. Oxf. Bull. Econ. Stat. 66, 425–433.
Gleser, L.J., 1973. On a theory of intersection-union tests. Inst. Math. Stat. Bull. 2, 2330.
Good, P., 2000. Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses, second ed. Springer Series in Statistics. Springer, Heidelberg.
Guerrero, V.M., 1987. A note on the estimation of Atkinson's index of inequality. Econ. Lett. 25, 379–384.
Hadar, J., Russell, W.R., 1969. Rules for ordering uncertain prospects. Am. Econ. Rev. 79, 25–34.
Hajargasht, P., Griffiths, W.E., Brice, J., Rao, D.S.P., Chotikapanich, D., 2012. Inference for income distributions using grouped data. J. Bus. Econ. Stat. 30, 563–575.
Hall, P., Racine, J., Li, Q., 2004. Cross-validation and the estimation of conditional probability densities. J. Am. Stat. Assoc. 99, 1015–1026.
Hampel, F.R., 1971. A general qualitative definition of robustness. Ann. Math. Stat. 42, 1887–1896.
Hampel, F.R., 1974. The influence curve and its role in robust estimation. J. Am. Stat. Assoc. 69, 383–393.
Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A., 1986. Robust Statistics: The Approach Based on Influence Functions. John Wiley, New York.
Hasegawa, H., Kozumi, H., 2003. Estimation of Lorenz curves: a Bayesian nonparametric approach. J. Econ. 115, 277–291.
Howes, S.R., Lanjouw, J.O., 1998. Poverty comparisons and household survey design. Rev. Income Wealth 44, 99–108.
Huber, P.J., 1981. Robust Statistics. John Wiley, New York.
Jenkins, S.P., 1995. Did the middle class shrink during the 1980s? UK evidence from kernel density estimates. Econ. Lett. 49, 407–413.

Jenkins, S.P., 2009. Distributionally-sensitive inequality indices and the GB2 income distribution. Rev. Income Wealth 55, 392–398.
Jenkins, S.P., Lambert, P.J., 1997. Three 'I's of poverty curves, with an analysis of UK poverty trends. Oxf. Econ. Pap. 49, 317–327.
Jenkins, S.P., Van Kerm, P., 2005. Accounting for income distribution trends: a density function decomposition approach. J. Econ. Inequal. 3, 43–61.
Jenkins, S.P., Burkhauser, R.V., Feng, S., Larrimore, J., 2011. Measuring inequality using censored data: a multiple imputation approach. J. R. Stat. Soc. Ser. A 174 (866), 63–81.
Johnson, N.L., Kotz, S., Balakrishnan, N., 1994. Continuous Univariate Distributions, second ed. John Wiley Series in Probability and Mathematical Statistics, vol. 1. John Wiley, New York.
Kakwani, N., 1993. Statistical inference in the measurement of poverty. Rev. Econ. Stat. 75, 632–639.
Kaur, A., Prakasa Rao, B.L.S., Singh, H., 1994. Testing for second order stochastic dominance of two distributions. Econ. Theory 10, 849–866.
Kleiber, C., Kotz, S., 2003. Statistical Size Distributions in Economics and Actuarial Sciences. John Wiley, Hoboken, NJ.
Kovacevic, M.S., Binder, D.A., 1997. Variance estimation for measures of income inequality and polarization. J. Off. Stat. 13, 41–58.
Langel, M., Tillé, Y., 2013. Variance estimation of the Gini index: revisiting a result several times published. J. R. Stat. Soc. A 176, 521–540.
Lemieux, T., 2006. Increasing residual wage inequality: composition effects, noisy data, or rising demand for skill? Am. Econ. Rev. 96, 461–498.
Li, J., Racine, J.S., 2006. Nonparametric Econometrics. Princeton University Press, Princeton, NJ.
Linton, O., Maasoumi, E., Whang, Y.-J., 2005. Consistent testing for stochastic dominance under general sampling schemes. Rev. Econ. Stud. 72, 735–765.
Lubrano, M., Ndoye, A.A.J., 2011. Inequality Decomposition Using the Gibbs Output of a Mixture of Lognormal Distributions, Working paper, Greqam 2011-19.
Maasoumi, E., Heshmati, A., 2008. Evaluating dominance ranking of PSID incomes by various household attributes. In: Betti, G., Lemmi, A. (Eds.), Advances on Income Inequality and Concentration Measures. Routledge Frontiers of Political Economy. Oxon, New York, pp. 47–69 (Chapter 4).
Majumder, A., Chakravarty, S.R., 1990. Distribution of personal income: development of a new model and its application to U.S. income data. J. Appl. Econ. 5, 189–196.
Mandelbrot, B., 1960. The Pareto-Lévy law and the distribution of income. Int. Econ. Rev. 1 (2), 79–106.
Marron, J.S., Schmitz, H.P., 1992. Simultaneous density estimation of several income distributions. Econ. Theory 8, 476–488.
Marron, J.S., Wand, M.P., 1992. Exact mean integrated squared error. Ann. Stat. 20, 712–736.
McDonald, J.B., 1984. Some generalized functions for the size distribution of income. Econometrica 52, 647–664.
McDonald, J.B., Xu, Y.J., 1995. A generalization of the beta distribution with applications. J. Econ. 66, 133–152.
McFadden, D., 1989. Testing for stochastic dominance. In: Fomby, T.B., Seo, T.K. (Eds.), Studies in the Economics of Uncertainty. Springer-Verlag, New York, pp. 113–134.
McLachlan, G.J., Peel, D., 2000. Finite Mixture Models. Wiley, New York.
Mills, J.A., Zandvakili, S., 1997. Statistical inference via bootstrapping for measures of inequality. J. Appl. Econ. 12, 133–150.
Modarres, R., Gastwirth, J.L., 2006. A cautionary note on estimating the standard error of the Gini index of inequality. Oxf. Bull. Econ. Stat. 68, 385–390.
Monti, A.C., 1991. The study of the Gini concentration ratio by means of the influence function. Statistica 51, 561–577.
Nygård, F., Sandström, A., 1985. Estimating Gini and entropy inequality parameters. J. Off. Stat. 1, 399–412.
O'Brien, G.L., 1984. Stochastic dominance and moment inequalities. Math. Oper. Res. 9, 475–477.
O'Brien, G.L., Scarsini, M., 1991. Multivariate stochastic dominance and moments. Math. Oper. Res. 16, 382–389.

Ogwang, T., 2000. A convenient method of computing the Gini index and its standard error. Oxf. Bull. Econ. Stat. 62, 123–129.
Osberg, L., Xu, K., 2000. International comparisons of poverty intensity: index decomposition and bootstrap inference. J. Hum. Resour. 35, 51–81, Errata. *Journal of Human Resources*, 35(3), inside front cover page, 2000.
Paap, R., Van Dijk, H.K., 1998. Distribution and mobility of wealth of nations. Eur. Econ. Rev. 42, 1269–1293.
Pareto, V., 1895. La legge della domanda. G. Econ. Ann. Econ. 10, 59–68.
Paul, S., 1999. The population sub-group income effects on inequality: analytical framework and an empirical illustration. Econ. Rec. 75 (229), 149–155.
Pen, J., 1974. Income Distribution, second ed. Allen Lane, The Penguin Press, London, pp. 48–59, Chapter 3.
Pittau, M.G., Zelli, R., 2004. Testing for changes in the shape of income distribution: Italian evidence in the 1990s from kernel density estimates. Empir. Econ. 29, 415–430.
Pittau, M.G., Zelli, R., 2006. Empirical evidence of income dynamics across EU regions. J. Appl. Econ. 21, 605–628.
Pittau, M.G., Zelli, R., Johnson, P.A., 2010. Mixture models, convergence clubs and polarization. Rev. Income Wealth 56, 102–122.
Polivka, A.E., 1998. Using Earnings Data for the Current Population Survey after the Redesign. Working Paper 306, US Bureau of Labour Statistics, Washington, DC.
Quah, D.T., 1997. Empirics for growth and distribution: stratification, polarization, and convergence clubs. J. Econ. Growth 2, 27–59.
Quirk, J.D., Saposnik, R., 1962. Admissibility and measurable utility functions. Rev. Econ. Stud. 29, 140–146.
Rao, C.R., 1973. Linear Statistical Inference and Its Applications, second ed. John Wiley, New York.
Ray, S., Lindsay, B.G., 2005. The topography of multivariate normal mixtures. Ann. Stat. 33, 2042–2065.
Reeds, J.A., 1976. On the Definition of von Mises Functionals: Research Report S 44. Department of Statistics, Harvard University, Cambridge, MA.
Resnick, S.I., 1997. Heavy tail modeling and teletraffic data. Ann. Stat. 25, 1805–1849.
Resnick, S.I., 2007. Heavy-Tail Phenomena: Probabilistic and Statistical Modeling. Springer Series in Operations Research and Financial Engineering. Springer, New York.
Robert, C.P., Casella, G., 2005. Monte Carlo Statistical Methods. Springer, New York.
Rothe, C., 2010. Nonparametric estimation of distributional policy effects. J. Econ. 155, 56–70.
Roy, S.N., 1953. On a heuristic method of test construction and its use in multivariate analysis. Ann. Math. Stat. 24, 220–238.
Rudemo, M., 1982. Empirical choice of histograms and kernel density estimators. Scand. J. Stat. 9, 65–78.
Ruppert, D., Sheather, S.J., Wand, M.P., 1995. An effective bandwidth selector for local least squares regression. J. Am. Stat. Assoc. 90, 1257–1269.
Sala-i-Martin, X., 2006. The world distribution of income: falling poverty and . . . convergence, period. Q. J. Econ. 121, 351–397.
Salem, A.B.Z., Mount, T.D., 1974. A convenient descriptive model of income distribution: the Gamma density. Econometrica 42, 1115–1127.
Sandström, A., Wretman, J.H., Walden, B., 1988. Variance estimators of the Gini coefficient: probability sampling. J. Bus. Econ. Stat. 6, 113–120.
Sarabia, J.M., 2008. Parametric Lorenz curves: models and applications. In: Chotikapanich, D. (Ed.), Modeling Income Distributions and Lorenz Curves. Springer, New York, pp. 167–190, Chapter 9.
Schechtman, A., 1991. On estimating the asymptotic variance of a function of *U* statistics. Am. Stat. 45, 103–106.
Schluter, C., 2012. On the problem of inference for inequality measures for heavy-tailed distributions. Econom. J. 15, 125–153.
Schluter, C., Trede, M., 2002. Tails of Lorenz curves. J. Econ. 109, 151–166.
Schluter, C., van Garderen, K., 2009. Edgeworth expansions and normalizing transforms for inequality measures. J. Econ. 150, 16–29.

Schmid, F., Trede, M., 1996. Testing for first-order stochastic dominance: a new distribution-free test. Statistician 45, 371–380.

Scott, D.W., 1992. Multivariate Density Estimation: Theory, Practice, and Visualization. John Wiley & Sons, New York.

Sen, A.K., 1976. Poverty: an ordinal approach to measurement. Econometrica 44, 219–231.

Sheather, S.J., Jones, M.C., 1991. A reliable data-based bandwidth selection method for kernel density estimation. J. R. Stat. Soc. 53, 683–690.

Shorrocks, A.F., 1983. Ranking income distributions. Economica 50, 3–17.

Shorrocks, A.F., 1995. Revisiting the Sen poverty index. Econometrica 63, 1225–1230.

Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman and Hall, London.

Singh, S.K., Maddala, G.S., 1976. A function for the size distribution of income. Econometrica 44, 963–970.

Stacy, E.W., 1962. A generalization of the gamma distribution. Ann. Math. Stat. 33, 1187–1192.

Stark, O., Yitzhaki, S., 1988. Merging populations, stochastic dominance and lorenz curves. J. Popul. Econ. 1, 157–161.

Stephens, M.A., 1986. Tests based on EDF statistics. In: d'Agostino, R.B., Stephens, M.A. (Eds.), Goodness-of-Fit Techniques. Marcel Dekker, New York, pp. 97–193.

Stone, M., 1974. Cross-validatory choice and assessment of statistical predictions (with discussion). J. R. Stat. Soc. B 36, 111–147.

Summers, R., Heston, A.W., 1991. The Penn world table (mark 5): an expanded set of international comparisons 1950-1988. Q. J. Econ. 106, 327–368.

Tachibanaki, T., Suruga, T., Atoda, N., 1997. Estimations of income distribution parameters for individual observations by maximum likelihood method. J. Japan Stat. Soc. 27, 191–203.

Thistle, P.D., 1989. Ranking distributions with generalized Lorenz curves. South. Econ. J. 56, 1–12.

Thuysbaert, B., 2008. Inference for the measurement of poverty in the presence of a stochastic weighting variable. J. Econ. Inequal. 6, 33–55.

Titterington, D.M., Makov, U.E., Smith, A.F.M., 1985. Statistical Analysis of Finite Mixture Distributions. Wiley, New York.

Verma, V., Betti, G., 2011. Taylor linearization sampling errors and design effects for poverty measures and other complex statistics. J. Appl. Stat. 38, 1549–1576.

Victoria-Feser, M.-P., 1995. Robust methods for personal income distribution models with application to Dagum's model. In: Dagum, C., Lemmi, A. (Eds.), Income Distribution, Social Welfare, Inequality and Poverty. In: Research on Economic Inequality, 6, JAI Press, Greenwich, pp 225–239.

Victoria-Feser, M.-P., 2000. Robust methods for the analysis of income distribution, inequality and poverty. Int. Stat. Rev. 68, 277–293.

Victoria-Feser, M.-P., Ronchetti, E., 1994. Robust methods for personal income distribution models. Can. J. Stat. 22, 247–258.

Victoria-Feser, M.-P., Ronchetti, E., 1997. Robust estimation for grouped data. J. Am. Stat. Assoc. 92, 333–340.

Xu, K., Osberg, L., 1998. A distribution-free test for deprivation dominance. Econom. Rev. 17 (4), 415–429.

Xu, K., Osberg, L., 2002. The social welfare implications, decomposability, and geometry of the Sen family of poverty indices. Can. J. Econ. 35, 138–152.

Yitzhaki, S., Schechtman, E., 2013. The Gini Methodology: A Statistical Primer. Springer Series in Statistics, Springer, New York.

Young, A., 2011. The Gini Coefficient for a Mixture of Lognormal Population. Working Paper.

Zheng, B., 1999. Statistical inferences for testing marginal rank and (generalized) Lorenz dominances. South. Econ. J. 65, 557–570.

Zheng, B., 2001. Statistical inference for poverty measures with relative poverty lines. J. Econ. 101, 337–356.

Zheng, B., 2002. Testing Lorenz curves with non-simple random samples. Econometrica 70, 1235–1243.

Zheng, B., Cushing, B.J., 2001. Statistical inference for testing inequality indices with dependent samples. J. Econ. 101, 315–335.

CHAPTER 7

Long-Run Trends in the Distribution of Income and Wealth

Jesper Roine*, Daniel Waldenström†
*Stockholm Institute of Transition Economics, Stockholm School of Economics, Stockholm, Sweden
†Department of Economics, Uppsala University, Uppsala, Sweden

Contents

Handbook of Income Distribution, Volume 2A
ISSN 1574-0056, http://dx.doi.org/10.1016/B978-0-444-59428-0.00008-4

Abstract

This chapter reviews the long-run developments in the distribution of personal income and wealth. It also discusses suggested explanations for the observed patterns. We try to answer questions such as: What do we know, and how do we know, about the distribution of income and wealth over time? Are there common trends across countries or over the path of development? How do the facts relate to proposed theories about changes in inequality? We present the main inequality trends, in some cases starting as early as in the late eighteenth century, combining previous research with recent findings in the so-called top income literature and new evidence on wealth concentration. The picture that emerges shows that inequality was historically high almost everywhere at the beginning of the twentieth century. In some countries this situation was preceded by increasing concentration, but in most cases inequality seems to have been relatively constant at a high level in the nineteenth century. Over the twentieth century inequality decreased almost everywhere for the first 80 years, largely due to decreasing wealth concentration and decreasing capital incomes in the top of the distribution. Thereafter trends became more divergent across countries and also different across income and wealth distributions. Econometric evidence over the long run suggests that top shares increase in periods of above-average growth, whereas democracy and high marginal tax rates are associated with lower top shares.

Keywords

Income inequality, Income distribution, Wealth distribution, Economic history, Top incomes, Welfare state, Taxation

JEL Classification Codes

D31, H2, J3, N3

7.1. INTRODUCTION

This chapter gives an overview of the evidence on long-run trends in the distribution of personal income and wealth. Our focus will be on empirical estimates of the respective distributions, but we will also cover some ideas that aim to explain the observed patterns. The *long run* refers, at best, to the period from around 1750, that is, the time around the British industrial takeoff, and onward, but in most cases the time span begins in the early twentieth century. As a result of data availability most of the evidence is based on today's developed economies and as a result generalizations will tend to be skewed toward this set of countries. However, this is not to say that the results are relevant for rich countries only. In many cases the data coverage starts at the very beginning of industrialization of today's developed economies, and in addition data is also available for several developing nations.[1]

The kinds of questions we try to answer in this chapter are: What do we know (and how do we know) about the distribution of income and wealth over time? Are there common trends across countries or over the path of development? How do the facts relate to proposed theories about changes in inequality? We will mainly draw on the advances made in the field over the past decade, but before outlining the contents of the chapter and its limitations, we want to recall some points in the development of evidence on long-run inequality leading up to the recent research.[2]

7.1.1 From the Kuznets Series, to Household Surveys, and Back Again

In the very beginning of his famous presidential address to the American Economic Association in 1954, Simon Kuznets outlined some ideals concerning the data required to study long-term changes in inequality (Kuznets, 1955). The "economist's pipe dream" that he described roughly corresponds to what we today would call a detailed, individual panel data set, preferably spanning several generations. He emphasized things such as being able to adjust incomes for household size, to capture "all units in a country rather than a segment either at the upper or lower tail," the importance of being able to control for transitory income fluctuations, being able to calculate individual life time incomes, and so on. He also stressed the importance of the relation between income and wealth (savings) for understanding the distributional dynamics over time.

[1] For a comprehensive treatment of developing countries, see Chapter 9 in this Handbook by Alvaredo and Gasparini (2014).

[2] In the *Handbook of Income Distribution*, Volume 1, two chapters provided a historical perspective; one by Peter Lindert on inequality in Britain and America starting with estimates for England and Wales as far back as 1688 (Lindert, 2000), and one by Christian Morrison on developments in selected European countries with observations as far back as the 18th century for France and Sweden (Morrisson, 2000). We naturally build on their chapters and focus on new findings extending the picture given in their respective chapters.

In many ways the development of inequality data for a long time after Kuznets' well-known speech focused on this "wish list." Even though important advances were, of course, made with respect to historical data, it is fair to say that the focus was on the construction of contemporary national household surveys and individual micropanel data sets.[3] Eventually much effort also went into making such data comparable across countries in projects such as the Luxembourg Income Study (LIS) and its more recent companion the Luxembourg Wealth Study. Building on these and other similar projects, compilations of data such as the World Income Inequality Database (WIID) have also been put together.[4] This development has indeed changed empirical inequality research for the better and made it possible to address a number of new and important questions. But the relative focus on microdata shifted attention away from some issues, and in particular questions regarding long-run developments. Given the relatively recent nature of most household survey data and microdata in general, "the long run" based to these sources naturally becomes quite limited, typically not covering more than the last couple of decades.[5] Such a relatively short time span is unfortunate because several issues concerned with economic development and structural change require a much longer time horizon.

However, recent research has changed things dramatically. Starting with the path-breaking work of Piketty (2001a, 2003), which extends the methods first used in the seminal work by Kuznets (1953) to generate a series of top income shares spanning the entire twentieth century in France using income tax data, similar efforts have followed for many countries. Using similar data and methodology, aiming at making estimates as homogenous as possible, new data are to date available for 26 countries. Most of these were collected in two volumes edited by Atkinson and Piketty (2007, 2010) that also contain chapters on methodological issues and summaries.[6] The full database is available online at http://g-mond.parisschoolofeconomics.eu/topincomes/, and as more studies are conducted, data is added here.

Most of the series span the whole of the twentieth century, sometimes even longer, making the resulting data set unique in its ability to address long-run issues. There are also

[3] Kuznets (1963) updated his series and added data for some more years and countries. Others, like Adelman and Taft Morris (1973), did compile an early data set on the distribution of income as far back as the late nineteenth century, but the reliability of this data was debated (e.g., Paukert, 1973). Contributions on the long-run development of wealth inequality include the comprehensive work by Atkinson and Harrison (1978) and studies by, for example, Lampman (1962), Atkinson et al. (1989), Soltow (1968, 1971), Lindert (1986, 1987), Wolff (1987), and Wolff and Marley (1989). Important contributions to the study of long-run income inequality include Soltow (1968, 1969, 1971), Williamson and Lindert (1980a,b, 1981), and Willliamson (1980). See the introductory chapter in Brenner et al. (1991) for references and an overview. See also further references in Lindert (2000) and Morrison (2000).

[4] Despite such efforts there are many remaining issues when using these data. Atkinson and Brandolini (2001) outline the pitfalls in using compilations of inequality data from different sources.

[5] See the introduction by Thomas Piketty in Atkinson and Piketty (2007).

[6] See also Piketty (2014) for an overview of this and related literatures.

other features of the data such as their relatively high frequency (often yearly), the possibility to decompose income by source, and the possibilities to study changes within the top of the distribution that have proven to be of importance and, as we shall discuss in more detail later, have led to new insights about inequality developments over the long run. This renewed interest in the long run and the reevaluation of historical sources has also led to new studies on the historical trends in the wealth distribution (e.g., Dell et al., 2007; Kopczuk and Saez, 2004; Piketty et al., 2006; see further Section 7.3).

The body of work extending and generalizing Kuznets' pioneering research is often labeled according to its focus on the top of the distribution. The "top income literature" is of course a correct description in the sense that it is based on observing only high-income fractions of the population (typically roughly the top decile and sometimes an even smaller share) and then relating their incomes to estimates of total income. From this it follows that top income shares cannot say anything about changes *within* a large share of the total population. But it does not follow that this data is only about the rich. As we will outline in more detail later, the top income literature is a contribution to both our understanding of long-run changes in overall inequality, as well as a more detailed understanding of developments within the top. Both aspects are important.

Finally, one should remember that it is not always a matter of *choosing* the right inequality measure for the question at hand. In fact, when it comes to the study of long-run inequality, the availability of *any data at all* is often the binding constraint. In such a situation the relationship between different measures becomes important, and we want to know things like: "What are the relationships between different inequality measures?" and "To what extent can this measure serve as a proxy for what we would ideally like to observe?" In the end, the approach to what we know and can know about inequality over the long run will have to be pragmatic. Such an approach calls for cautious interpretation, but not for resignation. We believe, using the words of Kuznets (1955, p. 4), that even "if the trends in the income structure can be discerned but dimly," we should continue to improve on our informed guesses. This is far from saying that the best we can do is to patch together scattered observations over time, using different sources and methods. In fact, many recent insights points exactly to the opposite. In the end we need to combine an understanding about what we are, in fact, observing, how different measures relate to each other, and an understanding of how they relate to the model or theory we wish to test.

7.1.2 Outline of the Chapter

This chapter has three parts in addition to this introduction: one on the trends in long-run income inequality, a second on trends in long-run wealth inequality, and a third on potential explanations of these trends and how they relate to some of the theories about what determines inequality.

7.1.2.1 Top Income Shares and Other Measures of Long-Run Income Inequality

In Section 7.2 we focus mainly on the new evidence on long-run income inequality that has come out of the top incomes project, including some new data points.[7] This means that income inequality is generally in terms of total income, that is, income from all sources, before taxes and most transfers. We briefly discuss the methodology and type of data used in this literature and then give an overview of the most important findings. First, we review the broad trends and to what extent the developments can be described as common for different groups of countries.[8] Second, we stress the importance of studying different parts within the top decile, as it turns out that this is a very heterogeneous group. Here we also present so-called shares-within-shares measures capturing the relative development between various top groups. Third, we emphasize the importance of decomposing income with respect to source of income. This is an aspect that has not received much attention in the past literature on historical inequality, but which can now be studied in more detail thanks to the nature of the income tax-based sources, and which turns out to be of great importance for the interpretation of inequality developments. We also discuss the importance of how to treat realized capital gains.

Thereafter we move on to relating the results based on top income shares with results based on other sources and measures of inequality. We consider both top share measures using somewhat different sources and methods, as well as other estimates of historical inequality based on other measures (wage dispersion across occupations, factor price differentials and differences in life prospects). In particular, we discuss and update the evidence on the issue of how good a proxy top income shares are for other measures of inequality. Putting everything together, we attempt to summarize the overall picture of income inequality developments for the period 1750–2010.

7.1.2.2 Long-Run Trends in the Wealth Distribution

In Section 7.3 we present the evidence on long-run developments of wealth inequality. Similar to the discussion of income inequality trends, we begin by reviewing the different data sources and empirical methods used to estimate the distribution of wealth over time. Much of the methodology used to study wealth distribution resembles the one used to examine trends in the income distribution. In particular, we often rely on top shares of a

[7] Using newly found statistical sources, we have calculated top income shares for Finland in the years 1865, 1871, 1877, and 1881.

[8] Most individual country studies are collected in the two volumes (Atkinson and Piketty, 2007, 2010), and all data are also available in the World Top Income Database (WTID). As new data becomes available for additional countries, this is added to the database together with information about sources and adjustments and so on. Taken together Atkinson and Piketty (2007, 2010) and the WTID provide details about particular aspects of data as well as the main suggested explanations for understanding individual country developments. Here we focus on developments that are joint for groups of countries or, possibly, for all countries.

consistently defined reference total population and their respective shares of an estimate of total wealth as our main measure of inequality. As in the case of top incomes, we also stress the importance of studying fractions within the top.

But there are also some important differences between studying income and wealth concentration. Personal wealth is more difficult both to define and to measure, and the nature of wealth data is also different from income data. Even though information on the distribution of wealth has been collected throughout history (the Doomsday Book from 1086 in England being an early and well-known example), wealth holdings have not typically been taxed directly in a systematic way. Assets have instead mostly been taxed on their transfer and in particular at the time of death. Indeed, most of the information we have on the distribution of distribution comes from inheritance or estate tax data, sometimes complemented by wealth data collected in connection to population surveys. The section describes how researchers have handled these challenges in estimating the wealth distribution and to what extent meaningful cross-country comparisons can be made.

After having discussed methodology, we move on to presenting the broad results emerging from this work covering 10 of today's industrialized economies from their respective eras of industrialization until the present. For a few countries (Finland, the Netherlands, Norway, and Sweden) the chapter also presents some new estimates of wealth concentration.[9]

7.1.2.3 Searching for Explanations

In Section 7.4, we then discuss the possible explanations behind the observed facts. How should we relate the shifts in the income and wealth distributions over time to other developments in society? To what extent are there global forces and events that affect all countries in similar ways (possibly with some time lag between countries)? What theories can shed light on shifts in capital incomes, what theories could explain increasing top wages? How should we think about the development of total income stemming from both wages and capital? What evidence do we have from regression analysis?

We begin by discussing some broad topics often suggested as a cause (and sometimes consequence) of inequality and sketch how the developments of these relate to our evidence. In particular, we will look at how our series correspond to broad global developments such as globalization, technological revolutions, wars and economic shocks, and patterns of economic growth. We then focus on more some specific aspects. First, we look at theories emphasizing capital incomes and also the interactions between earned income and capital as well as the cumulative effects of taxation. These things were all of key importance for the decline of top shares in the first half of the twentieth century and for the lack of recovery after the wars. We then consider some mechanisms that have

[9] We add observations of top wealth shares (to those already presented in previous studies) for Finland during 1987–2005, the Netherlands for 1993–2000 and 2006–2011, Norway in the 2000s, and Sweden in 2007.

been suggested to explain increased top wages such as skill-biased technological change, the rise in executive pay and related so-called super-star theories. These have all been suggested as important factors in the recent rise in top shares in many countries. Finally, we review some insights from econometric studies trying to use the new long-run inequality data to shed light on the developments.

Clearly our coverage of possible theories will be both selective and incomplete. In the end it is based on our subjective reading of which aspects we think are key for understanding the long-run developments of inequality, especially in light of the new evidence produced in the past decade. Furthermore, much of what we write about has been covered in previous overviews and surveys of the top incomes literature (Atkinson and Piketty, 2007, 2010; Atkinson et al., 2010, 2011; Leigh, 2009; Piketty, 2005; Piketty and Saez, 2006), overviews of the changing earnings distribution (Atkinson, 2008a) and in overviews on wealth concentration trends (Atkinson, 2008b; Davies and Shorrocks, 2000; Ohlsson et al., 2008; Wolff, 1996). In general, our aim is to focus on the most recent work in the field building on previous surveys such as Lindert (2000) and Morrison (2000).

7.1.3 What Is This Chapter Not About?

There is a lot of work and several issues regarding inequality over the long run that this chapter does not cover. As we see it, there are four major themes that we do not address but that are still closely related to what we discuss. Two of these omitted themes concern the descriptive scope of our chapter, whereas the other two relate more to the attempts to understand the developments.

First, we will not deal with issues of mobility but instead focus on repeated cross-sections of data.[10] A distribution where individuals constantly move in and out of the top (or bottom) of the distribution and where an individual's position 1 year says nothing of his or her position the next year is clearly very different to one where every individual keeps his or her place over time. Reality is obviously characterized by something in between the two extremes, but importantly the few studies that have been able to directly address this question (or aspects of it) conclude that trends in cross-sectional data are not driven by changes in mobility and do capture actual inequality.[11] In short, even if repeated cross sections of inequality, in theory, could be misleading when discussing changes in inequality over time, this does not seem to be the case in practice.

[10] For an extensive treatment of income mobility, see Chapter 10 by Markus Jäntti and Stephen Jenkins. For a more detailed study of mobility in the top of the income distribution, see Björklund et al. (2012).

[11] A few studies have looked at income mobility over the twentieth century, for example, Kopczuk et al. (2010) on U.S. intragenerational mobility and Björklund et al. (2009) and Lindahl et al. (2012) on Swedish intergenerational income mobility. See also Long and Ferrie (2007) on occupational mobility patterns in the United States and Great Britain since the nineteenth century.

Second, we will restrict our study in time to a period starting roughly at the beginning of the British Industrial Revolution (with data this far back being limited to a few data points for a small number of countries only), and with more comprehensive data starting in the beginning of the twentieth century. Recently there has been a lot of interesting work devising ingenious methods of estimating distributional outcomes in premodern societies.[12] All of this work certainly adds to our understanding of inequality in historical episodes as well as its long-run evolution. However, because these earlier figures are mostly based on occupational groupings or social class, we think one should be cautious when connecting our series to the estimates in earlier periods.

Third, we will not review theories about long-run inequality developments in any detail or with any attempt at fully coverage. We will instead outline some ideas and suggested mechanisms in a highly selective way to outline aspects that can help explain the key developments we find in the data.[13]

Fourth, we primarily discuss inequality as a left-hand side variable in an econometric sense. This means that our discussion will mainly be one about how we can understand the developments of inequality and its determinants and not so much about the consequences of inequality on other developments such as, for example, economic growth, political outcomes, or health.[14] Of course, such a distinction is somewhat artificial in the sense that the distribution of resources at any point forms the basis for economic and political decisions, resulting in outcomes that then create the distribution for the next period.[15] Many questions are, thus, ultimately not about one causing the other, but rather about the dynamic interplay over time. Nevertheless, it is often useful to separate questions in terms of how we think about the causality. In this separation we focus on understanding how and why inequality has changed, not on the consequences of inequality on other developments in society.

7.2. LONG-RUN TRENDS IN INCOME INEQUALITY

In his 1953 book *Shares of Upper Income Groups in Income and Savings*, Simon Kuznets produced the first comparable long-run income distribution series.[16] His main innovation

[12] See, for example, Soltow and van Zanden (1998), Milanovic (2006), Borgerhoff Mulder et al. (2009), Friesen and Scheidel (2009), and Milanovic et al. (2011).

[13] Chapter 5 in the *Handbook of Income Distribution*, Volume 1 (Piketty, 2000) provides an overview of theories of persistent inequality. See also Chapter 14 on inequality in macroeconomic theories and Chapter 15 on the relation between inheritance flows and inequality.

[14] See, for example, Atkinson (1997) and Aghion et al. (1999) for overviews of inequality and growth. Leigh et al. (2011) gives an overview of income and health.

[15] This is, for example, illustrated in the theoretical framework in Acemoglu et al. (2005).

[16] As Thomas Piketty (Piketty, 2007, p. 9) puts it: "These were the first long-run income distribution series ever produced (income distribution had been at the centre of speculative economic thought at least since the time of Ricardo and Marx, but few data were available)."

consisted in using U.S. income tax statistics over the period 1913–1948 and relating the incomes of those who paid taxes (the high-income earners) to an estimate of all personal income.[17] In his words:

> *The basic procedure is to compare the number and income of persons represented on federal income tax returns with the total population and its income receipts.[. . .] Since, except for a few recent years, tax returns cover only a small fraction of the total population—the fraction at the highest income levels—our estimates of income shares are only for a small upper sector. From the same source we can, with certain limitations, carry through the comparison for various types of income.*
>
> ***Kuznets (1953, p. xxix)***

The series for the United States, together with observations from England and Germany,[18] showing a secular decline of top income shares at least since the 1920s, formed the empirical basis of the famous "Kuznets curve" theory.[19]

Kuznets' series were not systematically updated, even if tax data and aggregate income sources of course continued to be available and developed.[20] In recent years, however, there has been what one may call a rediscovery of Kuznets' methodology and with it a significant increase in our knowledge about long-run changes in the distribution of income. Beginning with the influential work on long-run inequality in France by Thomas Piketty (Piketty, 2001a, 2001b, 2003) a number of researchers have created income inequality series using the same methodology for many countries (to date 26), and work is ongoing in many more.[21] For most countries the data spans the full length of the twentieth century, sometimes even longer. As Thomas Piketty phrases it in the introduction to the first of two volumes (Atkinson and Piketty, 2007, 2010) that collects

[17] Tax statistics had been used in several studies before but without relating them to top incomes. For example, see Bowley (1914) and Stamp (1914, 1916) for the United Kingdom and Crum (1935), Johnson (1935, 1937), and Tucker (1938) for the United States. In passing, it can be noted that a few years before Kuznets (1953) made his contributions, South African economists Herbert Frankel and Hans Herzfeld used a similar approach in a study of European income distribution in South Africa (Frankel and Herzfeld, 1943). Similarly, the Swedish economist Ragnar Bentzel independently published a study of the Swedish income distribution in the 1930s and 1940s, using almost the same approach as Kuznets did, that is, relying on historical tax returns data relating them to reference totals computed from national accounts (Bentzel, 1953).

[18] He also compared to some observations from India, Ceylon, Puerto Rico, Kenya, and Rhodesia, but in these cases there was no time series data.

[19] According to this, income inequality follows an inverse U-shape, rising with industrialization, as only a limited number of individuals initially work in the more productive sector, but then eventually declining, as the productive technology gradually spreads to the whole economy. One should note that he developed this idea because he saw his findings of decreasing inequality as a puzzle in face of other aspects that would work in the opposite direction. In particular, he stressed the cumulative effect of the concentration of savings that should increase inequality over time. We will return to this part of Kuznets' article and to the interplay between income and wealth in Section 7.4.

[20] Kuznets (1963) did return to the subject and added data for some more years and countries.

[21] Table 7.3 contains a list of countries and time spans for which we presently have data.

much of this work: "In a sense, all what we are doing in this project is to extend and generalize what Kuznets did in the early 1950s—except that we now have 50 more years of data and over 20 countries instead of one."

This—the long time span covered—is the most obvious advantage of the new data coming out of this project. For most countries the series start in the early 1900s and in some cases even further back. But there are other important aspects as well. First, data are typically high frequency (yearly), which has proven to be important for the interpretation of some historical developments, in particular the dramatic short-run shocks to top incomes in connection to the World Wars and the Great Depression. Second, the data offer a great deal of cross-country comparability as they are based on the same type of primary source across countries, income tax statistics, and there is typically no top coding of these data. Third, and perhaps most important, the data allow for a decomposition according to the source of income (i.e., earnings, capital income), which has proven to be of crucial importance for understanding long-run developments of inequality and, in particular, the interplay between income and wealth.

Naturally, there are important limitations with using these data as well. First of all, data are limited to the development of top income shares and do not reflect what happens in the rest of the distribution. (However, as we shall see in Section 7.2.3, it turns out that top income shares are highly correlated with more general distribution measures such as the Gini coefficient). Second, focus lies on pretax and transfer income. Third, the unit of analysis, as well as the income concept, is determined by the tax code, which differs both across countries and in some cases also over time within individual countries and means that we cannot make any adjustments for household size. It should be noted, however, that considerable effort has gone into adjusting for these changes to make country series at least consistent across time (but leaving some of the cross-country comparability problems unaddressed). Fourth, given the concerns in most countries with tax avoidance and tax evasion, tax statistics are potentially problematic as a source of information on incomes.

7.2.1 Methods and Data in the Top Income Literature

To answer the basic question, "What share of total income is received by some fraction of the population?" one needs to specify three things. First, we need to know what total income is, how it is defined, and how large it is. Second, we need to decide what population we are talking about (all individuals, all adults, all households, etc.). Third, we need information about the incomes of the subset of the population whose income share we want to relate to the total. The innovation of Kuznets (1953)—which was developed in Piketty (2001a) and has been the methodology used in the top-income literature—was to relate the assessed incomes of the taxpaying population to all household sector incomes. Because historically only those with the highest incomes were taxed and thus obliged to hand in personal tax returns, their incomes must be related to reference totals not only for everyone in the taxed population but also for the population as a whole. In other words, the reference total population and income need to also include individuals who did not file a tax return as well as their incomes. To construct these we must use aggregate sources such as population statistics (which is ample), census data (which do exist), and national accounts (which are scarce for historical eras). Top income shares can then be computed by dividing the number of tax units in the top, and their incomes, with the reference tax population and reference total income. Assuming that top incomes are approximately Pareto distributed, standard inter- and extrapolation techniques can be used to calculate the income shares for various top fractiles, such as the top 10% (P90–100) or the top 0.01% (P99.99–100).

In the following section, we will briefly outline the main issues associated with going from basic data to calculating homogenous income shares. This includes thinking about the nature of tax data and the typical adjustments made, the construction of a population total, the construction of an income total, the interpolation techniques used and the relation to shares-within-shares estimates, and finally some other issues such as part-year incomes. For a more detailed discussion on the methodology, see Atkinson (2007).

7.2.1.1 Tax Statistics and the Definition of Income

With the introduction of progressive income tax systems in many countries during the late nineteenth and early twentieth centuries came tabulations published by tax authorities over all income tax returns. These tabulations, often published annually, typically group incomes in different income brackets and, for each bracket, report the number of individuals (or, more generally, tax units) and the total income assessed. Table 7.1 exemplifies the type of information that is typically available in these tables with the case of Sweden in 1951.

As with most other income data sources, the tabulated income statistics does not correspond to any theoretically comprehensive definition of income but a definition determined by tax legislation.[22] And even more important, what is included in this tax income concept has often changed over time, and it varies across countries. To make estimates as comparable as possible, which has been a primary objective in each country study in the top income literature, one therefore needs to fix a definition for the income concept and then make adjustments to the tax data. The concept of income that has been used in almost all country studies of top incomes is some version of *total gross income*, defined as the sum of income from all sources, before taxes and transfers, but net of allowable deductions (mainly interest payments). Total gross income thus consists of factor income (labor earnings and capital income) plus occupational pensions, which equals market income, and in addition taxable transfer income (public pensions and some social

[22] The well-known Haig–Simons definition of income, for example, includes imputed rents, fringe employment benefits, and capital gains. These items are often not included in taxable income.

Table 7.1 Example of grouped income data from tax statistics: Sweden, 1951

Income class (tSEK)	Tax units	Income (tSEK)	Average income (tSEK)	Cumulative tax units	Cumulative tax units (%)	Cumulative income (tSEK)	Cumulative income (%)
0–0.6	154,414	43,002	0.3	3,969,635	100.00	23,274,169	100.00
0.6–1.0	222,940	111,491	0.5	3,815,221	96.11	23,231,167	99.82
1.0–1.5	235,230	261,731	1.1	3,592,281	90.49	23,119,676	99.34
1.5–2.0	239,850	392,751	1.6	3,357,051	84.57	22,857,945	98.21
2.0–2.5	225,110	503,851	2.2	3,117,201	78.53	22,465,194	96.52
2.5–3.0	193,550	552,984	2.9	2,892,091	72.86	21,961,343	94.36
3.0–3.5	189,590	591,231	3.1	2,698,541	67.98	21,408,359	91.98
3.5–4.0	177,800	682,637	3.8	2,508,951	63.20	20,817,128	89.44
4.0–4.5	180,030	761,374	4.2	2,331,151	58.72	20,134,491	86.51
4.5–5.0	182,160	917,150	5.0	2,151,121	54.19	19,373,117	83.24
5–6	373,140	2,144,387	5.7	1,968,961	49.60	18,455,967	79.30
6–7	385,710	2,633,731	6.8	1,595,821	40.20	16,311,580	70.08
7–8	345,720	2,753,591	8.0	1,210,111	30.48	13,677,849	58.77
8–10	437,440	4,096,471	9.4	864,391	21.78	10,924,258	46.94
10–12	177,860	1,927,328	10.8	426,951	10.76	6,827,787	29.34
12–15	112,370	1,507,572	13.4	249,091	6.27	4,900,459	21.06
15–20	72,140	1,216,108	16.9	136,721	3.44	3,392,887	14.58
20–30	43,010	1,005,136	23.4	64,581	1.63	2,176,779	9.35
30–50	14,958	621,526	41.6	21,571	0.54	1,171,643	5.03
50–100	5319	341,690	64.2	6613	0.17	550,117	2.36
100	1294	208,427	161.1	1294	0.03	208,427	0.90
Sum	3,969,365	23,274,169	5.9				

"tSEK" denotes thousand Swedish kronors, current prices.
Source: Statistics Sweden (1956, table 7).

benefits). Social Security contributions paid by employers and employees are generally excluded, as they are not part of the tax base.[23]

Even if the total gross income concept may seem like a clear enough definition, there are several broad categories of income that may cause problems of comparability both over time and across countries. One example is the tax treatment of transfers (often work-related such as sickness pay, unemployment insurance, and pensions) that are sometimes included in the tax base, for example, in the Nordic countries in recent decades. The reason to include them is that they are not viewed as "pure" transfers but rather part of a collective insurance scheme where you need to work in the first place to get the transfer.[24] Taxable transfers have typically become more important over time but are also very different in size across countries. Roine and Waldenström (2008) calculated top shares both including and excluding such transfers for Sweden. Their conclusion is that for most of the twentieth century the difference is small, but in recent years the increase in top income shares is notably larger for market income than for total income (including taxable transfers). In the year when the effect is the largest, the difference is almost 1 percentage point (about 15% of the income share), but it does not change the main trends though (and considering the importance of these systems in the Swedish context, this is likely to be an upper bound of the effect).

Another area is the inclusion (or exclusion) of capital income and, in particular, realized capital gains. Many countries have moved in the direction of excluding parts of capital income in their tax bases, and to the extent that such incomes accrue to top income groups, this would mean that top shares are underestimated over time. Although the income from interest-bearing bank deposits and corporate dividends are easily observed and included in most countries' taxable income concept, other capital incomes, such as the imputed rent of homeownership and realized capital gains, are more difficult to observe. Imputing income from owner-occupied housing requires information about housing stocks at the household level and has not been generally available over time. However, had it been possible to estimate homeownership rents, we believe that would have reinforced the equalization we observe over the twentieth century, possibly with a more ambiguous effect in the earlier period.[25] As for the impact of capital gains on long-run trends, this issue is discussed further in Section 7.2.2.3.

[23] Conceptually, including Social Security contributions in gross incomes could be motivated because studies have found that they are to a large extent ultimately born by workers through lower wages, which influences not only the estimation of income inequality (which becomes lower) but also the analysis of the redistributive effect of the tax-transfer system (see Bengtsson et al., 2012; Piketty and Saez, 2007).

[24] There may, of course, be other, for example, political economy, reasons for why politicians have decided to make these transfers taxable alongside factor income.

[25] In most developed economies home ownership spread throughout the population during the twentieth century as documented, for example, by Atkinson and Harrison (1978) for the United Kingdom and Waldenström (2014) for Sweden.

In many countries the historical income tax statistics also include information about the different sources of income, such as wage earnings, capital income, and business income, across the income distribution. In these tables, income earners are typically ranked according to total gross income, and then the amount of income from each source is listed within each gross income class. Table 7.2 displays an example of this kind of evidence for Sweden in 1951. Note that as in the case of total gross income, the reported incomes by source may not necessarily follow the theoretically most appropriate concepts but instead reflect definitions in the tax code. This is fairly clear in the Swedish 1951 example. The table consists of three, and perhaps even four, income sources reflecting capital income: interests and dividends (which are called "income from capital" in the tax data), (imputed) property income, realized capital gains, and the part of farm income adhering to imputed income from agricultural property. Also, what we would theoretically think of as labor income is not only contained in what is called "labor income" but also in business (or entrepreneurial) income as well as the part of farm income reflecting labor.

7.2.1.2 Reference Total for the Population

As tax statistics are based on the "tax unit" concept, the natural reference population is the total tax unit population of the country, had everyone filed a tax return. In countries like Australia, Canada, Italy, and Spain, taxes are filed individually, and the natural reference group is then the adult population defined as all residents above some age cutoff. In countries like France, Germany, and the United States, taxes are instead filed per family, which is typically defined as a married couple or a single adult (or an adult child living at home but with own income). In these cases the reference population becomes something like the adult population minus all married women (who file jointly with their husbands), with "adult" again defined as individuals above a certain age. In some countries, such as Finland, Sweden, and the United Kingdom, the definition of the tax unit has changed over time. In the Finnish case, for example, the family was the tax unit before 1935 when separate taxation of married couples was introduced. This was changed back again 1943, and the family was the tax unit again until the reintroduction of individual taxation in 1976. In Sweden the family was the tax unit before 1967 when a *choice* of filing individually was introduced. This was then the rule until individual taxation finally became compulsory in 1971. In the United Kingdom the family was the tax unit before 1991 when the system switched to being individual.[26] In all these cases the population total has to change accordingly. Sometimes there are overlapping periods or legislation that allows family or individual taxation. In these cases judgment has to be used to choose

[26] In addition to these legal changes there are cases where there is a discrepancy between how data are reported and the tax law. In Sweden, for example, data in the period 1951–1966 are reported individually even if couples were taxed jointly.

Table 7.2 Example of income by source: Sweden, 1951

Income class (tSEK)	Income from. . . (million SEK, share of total)						
	Labor	**Interests and dividends**	**Property**	**Realized capital gains**	**Farming**	**Business**	**Total gross income**
0–0.6	33 (75%)	6 (14%)	2 (4%)	1 (1%)	1 (1%)	1 (2%)	43 (100%)
0.6–1	99 (88%)	6 (5%)	2 (1%)	1 (0%)	2 (1%)	4 (3%)	111 (100%)
1–1.5	230 (87%)	15 (5%)	3 (1%)	1 (0%)	7 (2%)	7 (2%)	262 (100%)
1.5–2	339 (86%)	21 (5%)	5 (1%)	1 (0%)	17 (4%)	11 (2%)	393 (100%)
2–2.5	424 (83%)	22 (4%)	6 (1%)	2 (0%)	26 (5%)	25 (4%)	504 (100%)
2.5–3	470 (84%)	20 (3%)	7 (1%)	2 (0%)	36 (6%)	20 (3%)	553 (100%)
3–3.5	483 (81%)	16 (2%)	7 (1%)	2 (0%)	52 (8%)	33 (5%)	591 (100%)
3.5–4	546 (79%)	14 (2%)	4 (0%)	3 (0%)	79 (11%)	41 (5%)	683 (100%)
4–4.5	604 (79%)	13 (1%)	5 (0%)	1 (0%)	95 (12%)	45 (5%)	761 (100%)
4.5–5	750 (81%)	17 (1%)	2 (0%)	2 (0%)	95 (10%)	54 (5%)	917 (100%)
5–6	1772 (82%)	32 (1%)	7 (0%)	3 (0%)	221 (10%)	113 (5%)	2144 (100%)
6–7	2252 (85%)	27 (1%)	5 (0%)	3 (0%)	201 (7%)	150 (5%)	2634 (100%)
7–8	2403 (87%)	31 (1%)	4 (0%)	5 (0%)	181 (6%)	135 (4%)	2754 (100%)
8–10	3470 (84%)	56 (1%)	12 (0%)	10 (0%)	290 (7%)	270 (6%)	4096 (100%)
10–12	1550 (80%)	30 (1%)	13 (0%)	9 (0%)	167 (8%)	167 (8%)	1927 (100%)
12–15	1124 (73%)	31 (2%)	13 (0%)	13 (0%)	164 (10%)	177 (11%)	1508 (100%)
15–20	831 (67%)	53 (4%)	12 (1%)	12 (0%)	132 (10%)	187 (15%)	1216 (100%)
20–30	668 (65%)	48 (4%)	18 (1%)	14 (1%)	92 (9%)	180 (17%)	1005 (100%)
30–50	390 (61%)	40 (6%)	14 (2%)	11 (1%)	36 (5%)	142 (22%)	622 (100%)
50–100	207 (60%)	28 (8%)	12 (3%)	2 (0%)	16 (4%)	79 (22%)	342 (100%)
100	108 (52%)	29 (13%)	8 (3%)	1 (0%)	10 (4%)	53 (25%)	208 (100%)
Sum	18,753 (80%)	552 (2%)	162 (0%)	98 (0%)	1917 (8%)	1893 (8%)	23,274 (100%)

"tSEK" denotes thousand Swedish kronors, current prices.
Source: Statistics Sweden (1956, table 7).

the appropriate reference total. Table 7.3 gives an overview of the key features of the top income data used in all the 26 countries covered in our analysis, including the definition of tax units.

The main question based on these differences in tax unit concepts is, of course: "How important is this administrative setup and variation over time and across countries for the analysis of long run inequality trends?" Does it, for example, matter if we consider the population above 15 as adults or if we set the threshold to 20? Atkinson (2007) answered these questions under reasonable assumptions. The maximum difference between using an age cutoff at 15 instead of 20 (typically the alternative spans are smaller) results in a 6% (not percentage point) difference. If the top 1 percentile share were 10% with an age cutoff at 20, it would thus be 10.6% with a cutoff at 15. With respect to the effect of the tax unit being the individual or the family (or household), the maximum bounds are a little wider. Contrasting the extreme cases where top couples consist of individuals where either both earn the same, or one spouse has zero income, a top 1% share of 10% can be reduced to 8.3%, or increased to 11.8% when moving from joint to individual taxation. In cases where it has been possible to calculate top shares for individuals and couples, the difference is typically smaller. In general, Leigh (2005) showed that unless husbands and wives have equal income, individual-based data tend to (but must not) give rise to a more unequal income distribution than does the household-based data. Overall, the impact of changing tax units and age cutoffs are not likely to be important for the long-run trends we discuss below.

7.2.1.3 Reference Total for Income

When calculating the reference totals for income, there are basically two ways in which to proceed. Either one can start from the sum of all incomes reported on personal tax returns and then add items that are not included in the legal tax base as well as estimated incomes of individuals not filing taxes (not including children). Or one may start from the National Accounts item "Total Personal Sector Income" and from this broad concept deduct (estimates of) all items not included in the preferred definition of income. To the extent that data allow it, a calculation from both directions is of course desirable, as that would give an indication of the size of the potential difference between the respective procedures. In practice, these calculations may be difficult due to lack of data especially concerning early periods. In such cases the reference income total has typically been set to a share of GDP based on calibrations in periods when data are available (see Table 7.3 for an overview).

The following relation between the different parts (taken from Atkinson, 2007; Atkinson et al., 2011) is a useful illustration of the two procedures (starting either from the top and deducting items or from the bottom adding items),

Table 7.3 Key features in top income data

	Main source	Coverage	Tax unit, age cutoff	Income concept	Reference income basis	Capital gains included?
Australia	Atkinson and Leigh (2007a)	1921–2002 (82y)	Ind. 15+	GI	Nat. Acc.	Yes, where taxable
Argentina	Alvaredo (2010)	1932–1973 (39y)	Ind. 20+	GI	Nat. Acc.	No
Canada	Saez and Veall (2005)	1920–2000 (81y)	Ind. 20+	GI	Nat. Acc.	No (but reported after 1971)
China	Piketty and Qian (2009)	1986–2003 (18y)	Ind./Fam.	GI (incl. transfers)	Survey	No
Colombia	Alvaredo and Vélez (2013)	1993–2010 (18y)	Ind. 20+	GI	Tax stat.	Yes
Denmark	Atkinson and Søgaard (2013)	1870–2010 (97y)	Fam. 18+ -1969; Ind. 15+ 1970-	GI, AI.	Tax stat.	Yes
Finland	Jäntti et al. (2010)	1920–2004 (85y)	Ind. 16+	Gross/AI.	Tax stat.	No
France	Piketty (2001a, 2003)	1900–2006 (92y)	Fam.	GI	Nat. Acc.	No
Germany	Dell (2007, 2008)	1891–1918 (57y)	Fam. 21+	GI	Nat. Acc.	Yes, where taxable
India	Banerjee and Piketty (2009)	1922–1988 (71y)	Ind.	GI	Nat. Acc.	No
Indonesia	Leigh and van der Eng (2009)	1920–1939, 1982–2004 (34y)	Households	NI (excl. farm inc.)	Nat. Acc. -1939; Survey 1982-	No
Ireland	Nolan (2007)	1922–2000 (68y)	Fam. 18+	NI	Nat. Acc.	No
Italy	Alvaredo and Pisano (2010)	1974–2004 (29y)	Ind. 20+	GI excl. interest inc.	Nat. Acc.	No (but reported after 1981)
Japan	Moriguchi and Saez (2008)	1886–2005 (119y)	Ind. 20+	GI	Nat. Acc.	No
Mauritius	Atkinson (2011)	1933–2008 (74y)	Fam. 15+	GI (with adjustments)	Nat. Acc.	No
Netherlands	Salverda and Atkinson (2007), Atkinson and Salverda (2005)	1914–1999 (55y)	Fam. 15+	GI	Survey	No

New Zealand	Atkinson and Leigh (2007b)	1921–2002 (79y)	Fam. -1952; Ind. 1953-; 15+	AI. -1940; GI 1945-	Nat. Acc.	Yes, where taxable
Norway	Aaberge and Atkinson (2010)	1875–2006 (67y)	Ind. 16+	GI	Nat. Acc.	Yes
Portugal	Alvaredo (2010)	1936–2005 (64y)	Fam. 20+	GI	Nat. Acc.	No
Singapore	Atkinson (2010)	1947–2005 (57y)	Ind. 15+	GI	Nat. Acc.	No
Spain	Alvaredo and Saez (2009)	1933–2005 (49y)	Ind. 20+	GI	Nat. Acc.	No (but reported after 1981)
South Africa	Alvaredo and Atkinson (2011)	1913–2007 (71y)	Fam. 15+ -1989; Ind. 15+ 1990-	GI	Nat. Acc.	No (until 2002)
Sweden	Roine and Waldenström (2008)	1903–2006 (75y)	Fam. -1950; Ind. 1951-; 16+	GI	Nat. Acc. -1950; Tax stat. 1951-	Series with and without
Switzerland	Dell et al. (2007)	1933–1995/96 (31y)	Fam. 20+	GI	Nat. Acc.	No
United Kingdom	Atkinson (2005, 2007a)	1908–2005 (95y)	Fam. -1989; Ind. 1990-; aged 15+	GI	Nat. Acc. -1943; Tax stat. 1944-	Yes, where taxable
United States	Piketty and Saez (2003)	1913–2007 (96y)	Fam. 20+	GI	Nat. Acc. -1943; Tax stat. 1944-	No

Notes: See also Atkinson and Piketty (2007, 2010), Leigh (2009), Atkinson et al. (2011), and the World Top Income Database for information about the country statistics. Specifically, for several countries there are other studies that have contributed to the series. In column "Coverage," "y" denotes number of year observations. Tax units "Ind." and "Fam." stand for individual and family, respectively. "GI" denotes total income from all sources (labor, capital, business) gross of all deductions, whereas "NI" denotes income net of deductions. All incomes are before taxes and (most) transfers. "Nat. Acc." denotes that the reference income total is based on National Accounts data, typically some share of GDP or the sum of different aggregate income components, whereas "Tax stat." denotes that the reference total is derived from the sum of tax-assessed incomes plus some additional items (e.g., nonassessed income, imputed income from home ownership).

Total "Personal sector total income"
– Nonhousehold income (nonprofit institutions such as charities)
= Household sector total income
– Items not included in the tax base (such as employers' Social Security contributions, and nontaxable transfer payments)
= Household gross income reported to tax authorities
– Taxable income not declared by filers
– Taxable income of those who do not file tax returns
= Declared taxable income of filers

Using different reference totals can potentially have an important impact on the income shares. In their analysis of a number of alternatives for computing reference totals, some based on different National Accounts aggregates, some being fixed shares of GDP, and others departing from tax assessments adding estimated incomes of nonfilers, Roine and Waldenström (2010, Appendix C) show that there are indeed single years or episodes when differences are notable. Overall, however, the main trends in the results are robust to which alternative is chosen.

7.2.1.4 Interpolation Techniques and the Interpretation of the Pareto Coefficient

The historical income tax statistics typically come in the form of grouped observations, where income earners in different income brackets are separated by even income thresholds (see Table 7.2). The estimated top income shares, by contrast, present the share of total income earned by specific top fractions in the income distribution, such as the top 10, 5, 1, and 0.1 percentiles. These even fractions do almost never correspond to the even income thresholds observed in tax data. To get these top shares we therefore need to use interpolation and in some cases even extrapolation when shares of the highest top groups are estimated within the highest, open-ended income interval (see Table 7.2).

The most common interpolation procedure in the top income literature has been to assume that incomes in the top are Pareto distributed. This goes back to Pareto (1897), who was the first to make systematic observations of the size distribution of income. Given the nature of data, his observations were confined to the upper tail, and even though he initially thought that the Pareto function was a correct description of the whole distribution (with a bound at a "physiological minimum" >0), he eventually recognized that the distribution function over the whole population was probably hump-shaped and not Pareto distributed.[27]

[27] See, for example, Lydall (1968) for early references to the discussion of the shape of the top of the distribution and Atkinson (2007) with specific address to the top income literature. Some scholars have questioned the validity of the assertion that top incomes are Pareto distributed and instead applied other interpolation techniques, for example, mean-split histograms, to construct exact top shares (Atkinson, 2007; Atkinson et al., 2011).

The Pareto law says that incomes within the top of the distribution can be characterized by a power function of the form

$$f(y) = ky^{-\alpha}, \tag{7.1}$$

where y denotes income and k and α are constants. The parameter α in (7.1) is called "Pareto's alpha" or the "Pareto–Lorentz coefficient," and it reflects the degree of inequality, or the steepness of the income distribution; the higher α the lower the inequality. To see this, we can express the average income $\tilde{y}$among people earning above a certain "base" income b as a function of the α as

$$\tilde{y}= \left(\frac{\alpha}{\alpha - 1}\right) b. \tag{7.2}$$

That is, at any income level b, the average income above is $\left(\frac{\alpha}{\alpha-1}\right)$ times as large. As $\alpha \to \infty$ the difference between the level b and those above goes to zero, while as $\alpha \to 1$ the distribution moves toward everything being concentrated in the top. This economically more intuitive interpretation of $\left(\frac{\alpha}{\alpha-1}\right)$ has lead to this ratio sometimes being called the "inverted Pareto–Lorentz β coefficient," $\beta = \frac{\alpha}{\alpha-1}$. This coefficient provides a tractable association between a theoretical inequality index and the empirically estimated top income shares.[28]

The assumption of a Pareto distributed upper income tail has been confirmed by several studies using individual microdata for years when such data are available.[29] But again, the results coming out of the top income literature do not hinge on this assumption. Several studies of top income shares have instead of Pareto interpolation estimated top shares using slightly different techniques, primarily mean-split histograms (see table 4 in Atkinson et al., 2011, for details).

7.2.1.5 Tax Avoidance and Tax Evasion

Problems with tax avoidance and evasion are present in all studies of income inequality based on data from personal tax returns.[30] Importantly, though, overall underreporting does not necessarily change income *shares*. If incomes are missing in equal proportion across the distribution and are also missing from the reference total, the shares are unaffected. If, however, income is missing in equal proportions in tax statistics but not from the reference total (as could be the case if we combine tax statistics and National Accounts statistics) then we will underestimate top shares (and overestimate the share of the rest of the population) because we simply allocate the income not observed for the top earners as

[28] The characteristic of the β that it is constant within the income top, that is, it does not depend on the level of base income b, has been shown not to be empirically true for most countries, however.

[29] See Feenberg and Poterba (1993, 2000) for the United States, Piketty, 2001a,b for France, and Atkinson et al. (2011) for further examples.

[30] See Slemrod and Yitzhaki (2002) for an overview of tax evasion and avoidance and Slemrod (2000) for an overview of several issues concerning the economics of taxing the rich. We will not emphasize the distinction between legal tax avoidance and illegal tax evasion, as we are interested in all missing income. Seldon (1979) proposed the term "Tax Avoision" to capture the blurring between the two.

being received by the rest of the population. If avoidance is more important in the top, then we will of course also underestimate their share, whereas the impact of underreporting being more prevalent in the rest of the population typically creates a bias in the opposite direction, but it also depends on the construction of the reference total.

The main potential problem for assessing the trends, however, is the extent to which avoidance and evasion is very different across countries or changes in a systematic way over time. It could, for example, be argued that the increased tax rates seen over the twentieth century have given taxpayers increased incentives to avoid taxation. But this would be ignoring that the same increase in tax rates have given tax authorities increased incentives to collect taxes. Broadly speaking, high tax countries are also better at enforcement.[31] In the recent top income literature virtually all studies include sections on the issue of tax avoidance and evasion. Unsurprisingly, these all point to avoidance and evasion in various forms being present in all countries but the overall picture that emerges is that it is very unlikely that this would have a significant impact on the overall trend (see Atkinson et al., 2011, for details). To illustrate, Italy stands out as a country where evasion is much larger than other OECD countries but Alvaredo and Pisano (2010) still concluded that this does not change the main development of inequality. Dell et al. (2007) looked at the impact of assuming that *all* foreign income in Switzerland goes to French taxpayers and concluded that this would have a marginal effect on French top income shares. Similarly, Roine and Waldenström (2008) estimated the impact of capital flight from Sweden and concluded that even if the absolute numbers are sizable, and the impact on top income shares is nontrivial, the effect does not alter the general conclusions. Under the extreme assumption of attributing all unexplained residual capital flows out of Sweden since the 1980s to the top 1% income group, this increases their share by about 25%. This is significant, but it barely changes Sweden's rank or trajectory in relation to other countries.

The areas where avoidance and evasion responses are most likely to have a significant impact are on short-run fluctuations and when it comes to distinguishing the source of income. When ranking the importance of different behavioral responses to taxation Slemrod (1992, 1996) placed timing of economic transactions at the top as most responsive to tax incentives. Examples of this are clearly visible in the form of spikes in certain years, in particular when including realized capital gains (e.g., in connection to the tax reform act in the United States in 1986, in connection to changes in capital gains taxes in Sweden in 1991 and 1994, the year before the increased tax on dividends in Norway in 2006). As the second most important response to taxation Slemrod identified financial and accounting responses. This could take the form of income shifting between being corporate or personal, but also shifting the reported source of income. There are, for example, clear incentives for individuals to shift earnings to take the form of capital income in dual tax systems

[31] Overall, there is evidence that taxation is a key component of administrative capacity of government (Besley and Persson, 2009, 2013). See also Friedman et al. (2000).

where capital taxes are lower than wage taxes. Such income shifting does not lead to aggregate effects but may be of importance when interpreting shifts across income sources.

The issue of avoidance and evasion is clearly potentially important and should not be dismissed. Still, it is striking that not even in evaluating cases that we have reason to believe are among the more extreme do we see effects that dramatically change the overall trends. Also, as noted by Atkinson et al. (2011), the fact that some incomes (typically from capital) are tax exempt probably has a more important impact on inequality than underreporting.

7.2.1.6 Other Issues

In addition to the preceding, there are many other details in the historical income distribution data that call for attention and possibly correction. For example, in any given year individuals move into and out of the relevant tax unit population, some become "adults" due to age reasons, some die, some move into the country, others move out, some get married, others divorce. This mobility affects the relevant population, and it also creates "part-year incomes" that show up as low incomes in the data. Another potential difficulty is that tax years may not correspond to calendar years. Beside the problem of how to label observations, this may create problems if reference data are collected for calendar years (as is often the case). Fortunately these problems turn out not to be very large in quantitative terms.[32]

7.2.1.7 So Can We Trust the Top Income Data?

How should one deal with the challenges mentioned earlier that are associated with using historical income tax statistics? In past research scholars have suggested different approaches, including calculating theoretical bounds of the size of potential errors and employment of alternative sources that offer external checks of the order of magnitude by which an estimate could be wrong. In the end, however, one must make a number of judgment calls to select a final preferred series, and such calls can of course always be questioned. Having said that, considerable effort has gone into the construction of the series for each individual country with the explicit aim to make the series as homogenous as possible. We actually think that a hallmark of this research has been to take data quality issues very seriously and wherever possible produce estimates under different assumptions to be transparent about the effects of each individual choice made. In most cases where there are alternative ways to proceed, all alternatives have been explored and to the extent that this affects the results this is reported. The end result, we believe, is a data set with robust conclusions about the development of top income shares over time.

[32] For example, Atkinson (2007b) reported that part-year incomes reduced the top 10% income share by 0.3 percentage points in 1975–1976 (out of a total of about 25% in that year).

7.2.2 The Evidence and What We Learn

We identify three main themes in the empirical results. These themes form the basis for the three subsections that follow. The first addresses the overall evolution of income inequality as reflected in top income shares of the 26 countries covered here. The second theme is about the results showing a considerable heterogeneity among groups *within* the income top, especially differences in the top percentile and those in the lower part of the top decile. The third theme considers the role of decomposing total incomes by source, that is, assessing whether the recorded trends are, for example, driven by changes in the earnings distribution or whether they are based on shifts in the returns to personal wealth.

7.2.2.1 Common Trends or Separate Experiences?

Figure 7.1 illustrates the top 1% income share over the period 1870–2010 for all observations we have to date. Clearly this kind of illustration is not meant to be readable in the sense that the development of individual countries is discernible; rather it illustrates the extent to which there are truly common trends globally.[33]

The overall picture that emerges is one where the top 1% income share hovers around a relatively high level up until the First World War (in the few countries for which data exist), and then declines steadily over the twentieth century up until around 1980. After 1980 there seems to be a more scattered pattern. In some countries, in particular the United States and the United Kingdom, and in Anglo-Saxon countries more generally, top shares have increased significantly, whereas developments in other places, in particular in some Continental European countries, are close to flat after 1980.

In the literature on top income shares, much emphasis has been put on the diverging pattern between Anglo-Saxon countries and continental Europe.[34] As a result of the recent additions of new evidence from other countries, however, it is motivated to go beyond this dichotomy and incorporate the experiences of countries in other parts of the world.[35] We extend the division and examine inequality trends across six different country groups:

- Anglo-Saxon countries (Australia, Canada, Ireland, New Zealand, United Kingdom, and the United States)

[33] This question can also be asked in a more systematic way by identifying common trends and structural breaks in the series using econometric techniques; see Roine and Waldenström (2011).

[34] This difference is one of the main findings in the recent research on top incomes. Indeed, the title of the first of two volumes (Atkinson and Piketty 2007, 2010) collecting much of this work is *Top Incomes over the Twentieth Century: A Contrast Between European and English-Speaking Countries*.

[35] Alternatives to geography as a basis for country grouping exist. One is to divide them based on their participation in the Second World War. Another grouping could be based on types of "welfare state regimes," using the terminology of Esping-Andersen (1990). Here Japan fits into the corporatist tradition corresponding roughly to the Continental European countries, although that is perhaps most true for a more recent subperiod than for the whole of the twentieth century.

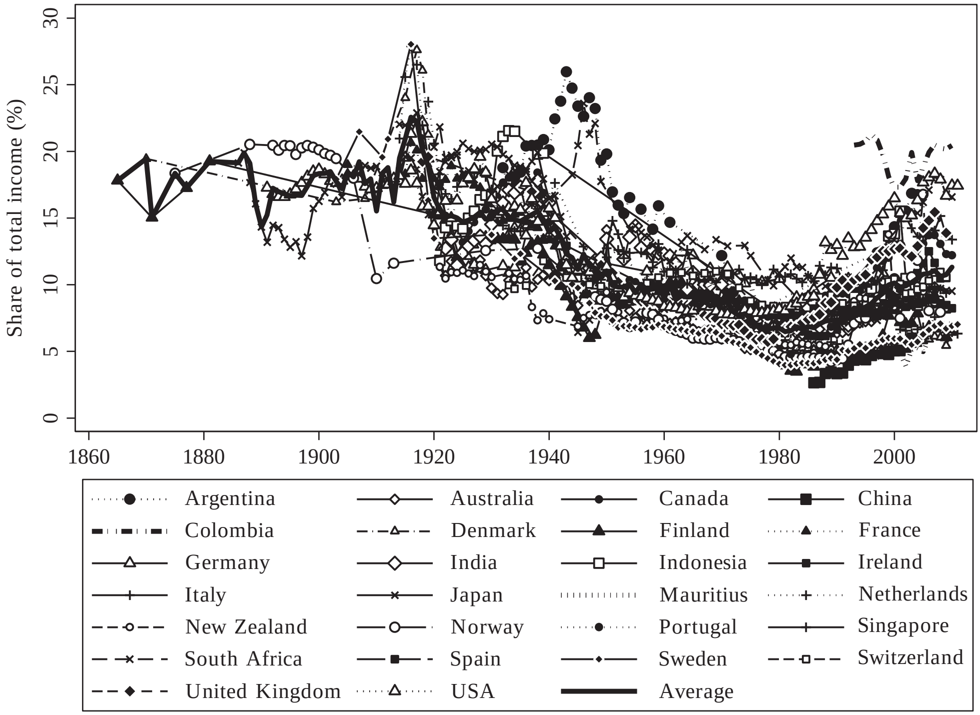


Figure 7.1 Top 1% income share in 26 countries, 1870–2010. *Source: See main text for description of the series and the World Top Income Database for sources.*

- Continental European countries (France, Germany, Italy, the Netherlands, Portugal, Spain, and Switzerland)
- Nordic countries (Denmark, Finland, Norway, and Sweden)
- Asian countries (China, India, Indonesia, Japan, and Singapore)
- African countries (Mauritius and South Africa)
- Latin American countries (Argentina and Colombia)

Figure 7.2 presents the long-run evolution of top income percentile shares in these six country groups.[36] Looking at the overall long-run development, there are clear similarities across the groups. They all exhibit a sharp decline in the top shares over the twentieth century, beginning around the time of the First World War and further reinforced by dramatic drops around the Second World War. Wartime shocks thus appear to have

[36] The creation of geographical country groups is problematic. Some of them are fairly homogenous, for example, whereas the other groups are more diverse, in particular the Asian group. In fact, apart from being Asian countries, it is hard to find a priori reasons for why they should constitute a group. Moreover, the small number of Latin American and African countries also pose problems in terms of representativeness.

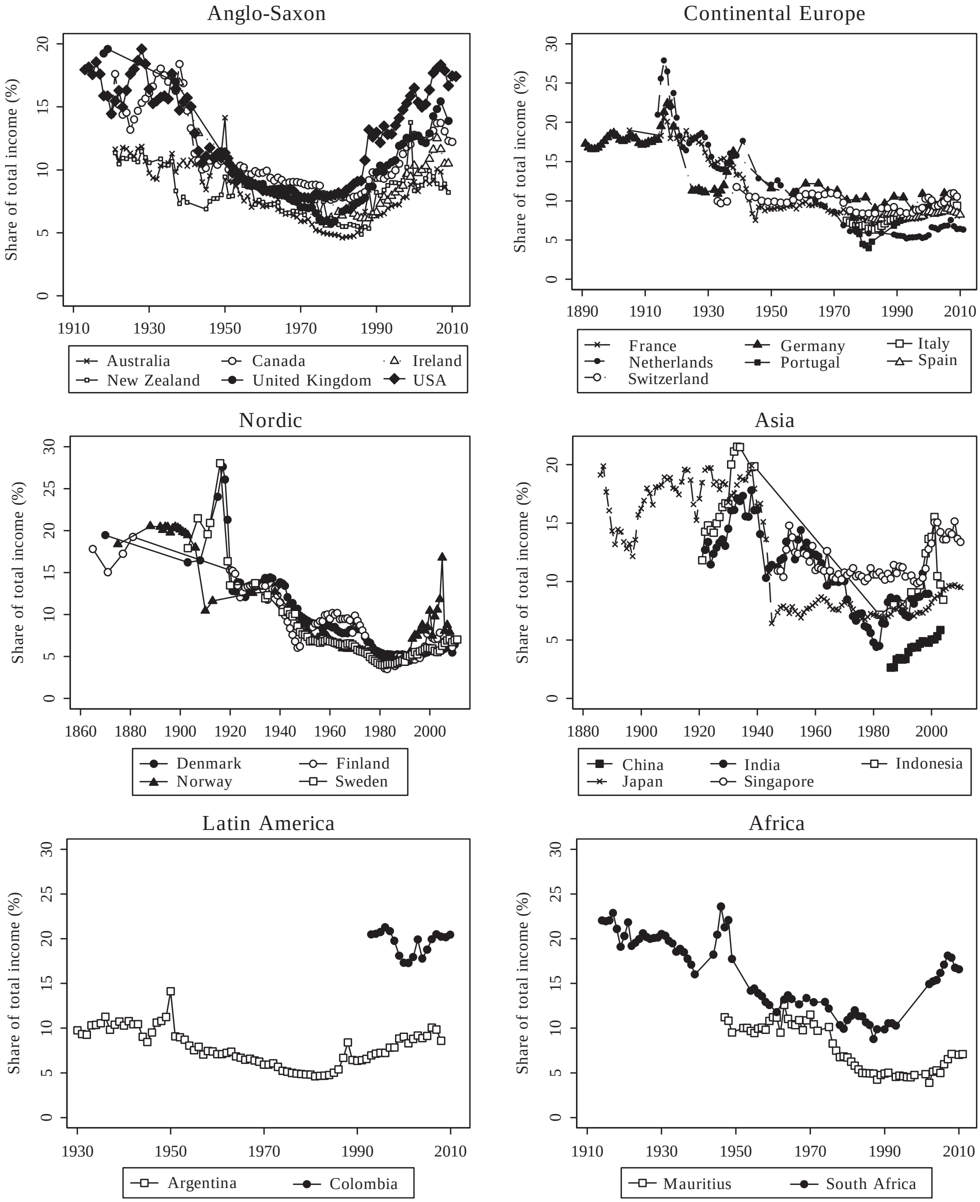


Figure 7.2 Top 1% across country groups. *Source: See Figure 7.1.*

had a large impact on top income shares. Everyone was probably affected by the wartime trade disruptions and new regulations of most goods and labor markets, but when it comes to specific surtaxes on wealth and high incomes or even the bombings of factories and similar capital destroying events, these were probably more important for the incomes of the rich. Having said this, the period 1914–1945 was also associated with periodic booms and asset price bubbles set off by a combination of highly expansionary fiscal policies and the economies being relatively closed. In both Denmark and Sweden, top income shares actually spiked in the midst of the First World War, and this is generally regarded as a consequence of the boom and asset price bubbles (Atkinson and Søgaard, 2013; Roine and Waldenström, 2008).

The twentieth century equalization trend in the top income shares continued up until the 1980s when it either flattened out in some countries or was reversed into increasing top income shares. That these common trends over the past century are in fact statistically significantly joint across countries has been shown recently by Roine and Waldenström (2011) in an analysis of common and country-specific trends and structural breaks in top income shares.

Notwithstanding the similarities, the evidence also indicates variation across countries within the geographical groupings reported earlier. For example, the upward trend in top income shares began in the late 1970s in the United States, Canada, and the United Kingdom, but started about 5–10 years later in Australia, New Zealand, and Ireland (though Ireland had a short-term peak around 1980). Within Continental Europe, most countries have not experienced stark increases in the top percentile share except for Portugal where it more than doubled between 1980 and 2000. The Asian data are not sufficiently complete to allow for conclusions about country differences: Japan and India appear to follow roughly similar patterns over time, with stable inequality levels before and after the dramatic shift in the 1940s when not only war but also profound institutional change hit these two countries. Since 1980 all the five Asian countries have exhibited an increasing top share. In Latin America and Africa, variation is small but so is the sample, and we cannot draw any conclusions from these results until we increase the number of observations.

Altogether, this analysis shows that with respect to the development of inequality, almost all countries display a secular decline in top income shares over the twentieth century up until around 1980. This decline is substantial: top percentile shares drop from around 20% of total personal income at the beginning of the 1900s to between 5% and 10% around 1980. In many countries much of this decline is concentrated around the World Wars and the Great Depression. Around 1980 the decline in top shares stopped, and in most countries they started to increase. This increase is substantial in Western English-speaking countries (Australia, Canada, New Zealand, the United Kingdom, and the United States) as well as in China and India. It is more modest but still clear in both some Nordic countries (Sweden, Finland, and Norway, but less clear in Denmark) and some Southern European countries (Italy and Portugal, but less clear in Spain), whereas finally, the development in some Continental European countries (France, Germany, the Netherlands, and Switzerland) and in Japan is close to flat.

7.2.2.2 The Importance of Developments Within the Top Decile

In income inequality research, top income earners are often defined as everyone in the top decile (P90–100) of the income distribution. However, recent studies following Piketty (2001a) have shown that the top decile is very heterogeneous.[37] For example, the income share of the bottom nine percentiles of the top decile (P90–99) has been remarkably stable over the past century in contrast to the share of the top percentile (P99–100), which fluctuated considerably. Moreover, although relatively high wage earners dominate in the lower group of the top decile, capital incomes are relatively more important to the top percentile. Figure 7.3 shows the development of the P90–99 income share over the period 1870–2010. Whereas the top percent income share fell by roughly a factor between 2 and 4 in the period until 1980 and has thereafter increased by a factor 2 in some countries, the long-run share of the P90–99 group has on average been relatively stable around 20–25% over the whole period.

An alternative way of studying income concentration is to express it in terms of the income share of certain top groups within the income share of another, larger, top group.

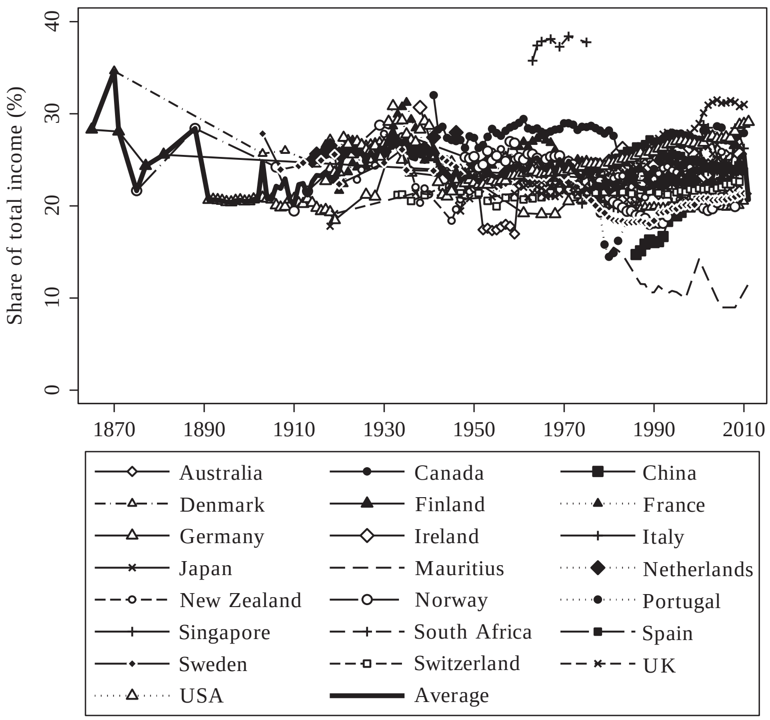


Figure 7.3 Trends in "next 9" percentiles in the top decile (P90–99), 26 countries. *Source: See Figure 7.1.*

[37] See Atkinson and Piketty (2007).

There are at least two merits with this approach. First, it measures the inequality within the top of the distribution, which is different from inequality overall especially when considering theories that predict a widening gap among the rich. Second, the top income shares may contain measurement error through the estimated reference total income held by the full population. By dividing the top income percentile by the top income decile (i.e., P99–100/P90–100), we get a "shares-within-shares" ratio that eliminates the reference total.[38]

Figure 7.4 shows the trend in the shares-within-shares ratio where we divide the top income percentile by the top income decile. It largely resembles the evolution seen in Figure 7.1, with a stable and relatively high level up to the 1910s and then a declining trend up until about 1980, after which an increase can be observed in some countries. This indicates both a degree of robustness of the overall trends in top income shares shown earlier and that concentration within the top has also changed over time.

Taken together, the evidence suggests that there are substantial differences in the long-run development between different groups in the top income decile. In fact, most

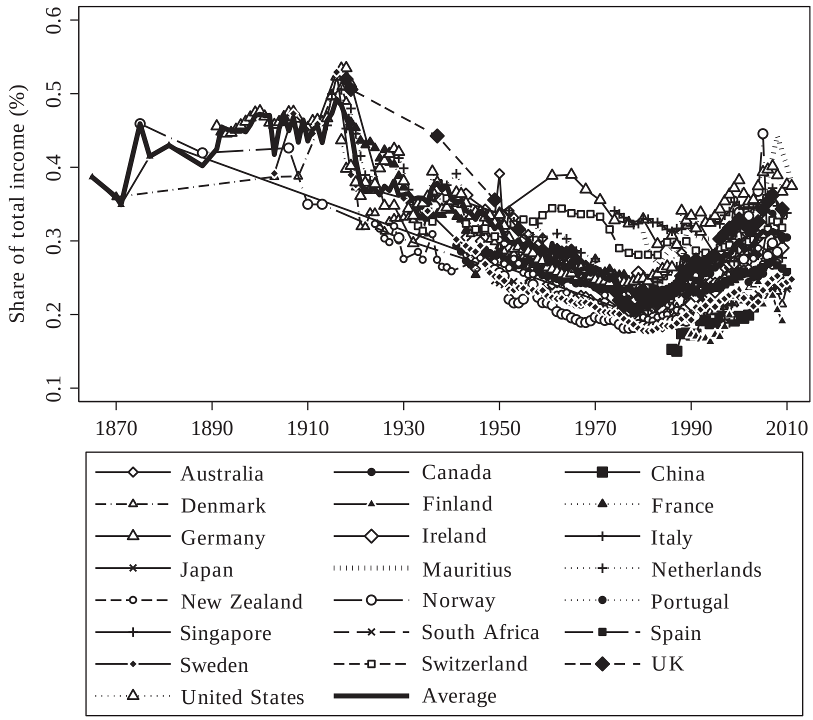


Figure 7.4 Shares-within-shares in top incomes (P99–100/P90–100). *Source: See Figure 7.1.*

[38] To see that this removes the influence of reference totals, note that P99–100 = $Y_{\mathrm{Top1}}/Y_{\mathrm{All}}$ (for income Y) and P90–100 = $Y_{\mathrm{Top10}}/Y_{\mathrm{All}}$. Hence, P99–100/P90–100 = $(Y_{\mathrm{Top1}}/Y_{\mathrm{All}})/(Y_{\mathrm{Top10}}/Y_{\mathrm{All}}) = Y_{\mathrm{Top1}}/Y_{\mathrm{Top10}}$.

of the observed overall changes in inequality are driven by decreasing or increasing shares of income earned by the top percentile group (P99–100), whereas the income share of the rest of the top decile in most countries is remarkably constant over the whole of the twentieth century.[39]

7.2.2.3 The Importance of Capital Incomes and Capital Gains

A major finding of the recent top income literature is that capital incomes are crucial for the development of income inequality over the long run (Atkinson and Piketty, 2007, 2010; Atkinson et al., 2011). Although wage earnings have always comprised the bulk of incomes among the masses, in the top of the distribution incomes have come from both labor and capital. As a consequence, the variation in top income shares can be expected to largely reflect changes in capital income flows. Some of these capital incomes are returns to corporate ownership, some are coupon yields on fixed-interest securities, whereas others come in the form of rental payments from tenants, interest earnings on bank deposit accounts, or as capital gains on financial or nonfinancial assets owned or sold. Our understanding of inequality trends over the long run requires that we closely examine the nature of these capital incomes and, in particular, the association between the distributions of income and personal wealth.

Unfortunately, few countries offer long-run distributional evidence by income source. Figure 7.5 shows the share of capital income (excluding capital gains) in total income since 1920 for the top percentile (P99–100) and the next nine percentiles in the top decile (P90–99) in four countries: Canada, France, Sweden, and the United States. Some notable results stand out. First, the importance of capital income clearly increases in the income level; in all cases capital is a more important source of income for the P99–100 than for the P90–99 group. Second, there was a sharp drop in the share of capital income around the Second World War, with the capital income share dropping by roughly half. This result clearly matches well with the findings of a similar drop in wealth concentration around the time of the war (see the following section for further information), whether due to wartime destruction or increased taxation and regulatory pressures.[40] Third, there is no clear uniform trend in recent decades; in the United States the importance of capital income seems to decrease, in France and Sweden the opposite appears true, while in Canada no clear trend is discernible.

For some countries, such as Sweden, the historical income tax statistics offer a possibility to cross-tabulate taxable wealth and income across both wealth and income distributions for most of the twentieth century (see Roine and Waldenström, 2008, 2009).

[39] This stability is even more marked when looking at the lower half of the top decile (P90–95). For example, this share moves around 9–11% in Sweden and between 10% and 13% in the United States over the entire twentieth century.

[40] It is interesting to note that this change in the role of capital is almost equally important in the case of Sweden, which did not take part in the war as in France and the United States.

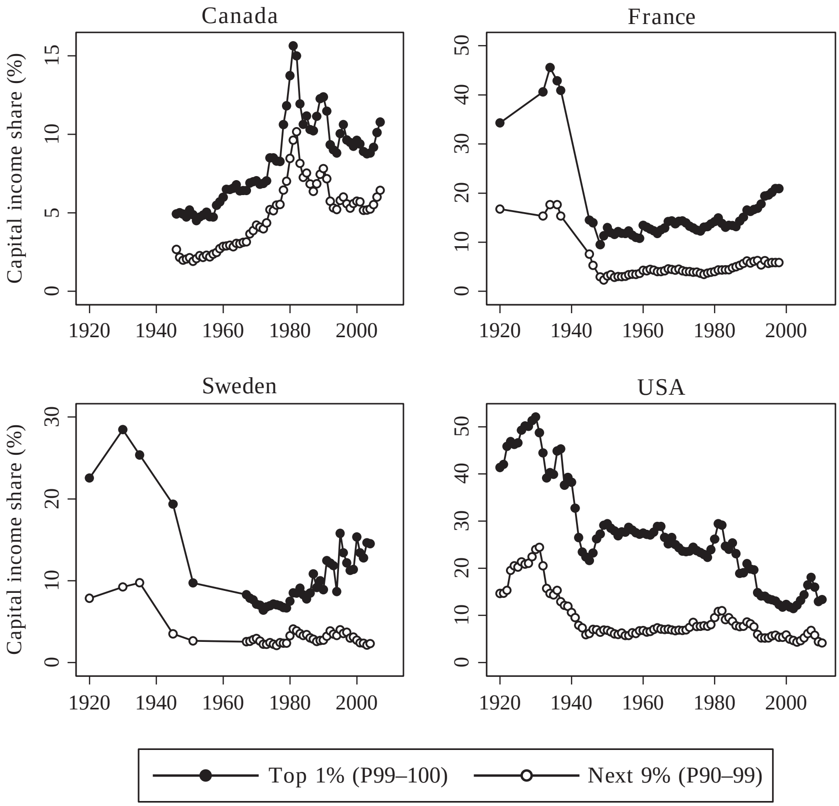


Figure 7.5 Capital income share in total gross income, 1920–2010 (%). *Source: Saez and Veall (2005), Veall (2010) for Canada, Piketty (2001b) for France, Roine and Waldenström (2008, 2010) for Sweden, and Piketty and Saez (2003, and updates) for the united States.*

Although not a complete data source, this allows us to get more insights into the interrelationship between income and wealth and how this matters for the long-run evolution of income inequality. The Swedish evidence indicates that the total wealth share held by people in the top income percentile decreased before 1950, in particular in the interwar period. By contrast, the "high-wage" income earners in the P90–95 income fractile increased their wealth share substantially over the same period, mainly in the 1910s and 1930s. The natural interpretation of these changes is that wealth as a source of income for the very rich declined in this period while, at the same time, moderately rich groups with high incomes accumulated new wealth. However, the drastic drops in Swedish capital income shares between 1930 and 1950 in the entire top decile seen in Figure 7.5 is not mirrored in their relative wealth share. Possibly this could be due to some wealth not being fully covered in taxable wealth because of definitions of the tax code or tax avoidance.

Capital gains turn out to be an additional important and interesting question. Theoretically capital gains, realized and unrealized, are undoubtedly a source of income in the classic Haig–Simons definition.[41] But in practice, capital gains represent a highly complicated income component to include in an individual's income. First, to the extent that they are observable at all, capital gains only appear on tax returns at the point of realization, making it difficult to properly allocate them in time. In many countries' tax codes (e.g., Spain and Sweden up until 1991) parts of the realized capital gains are tax exempt depending on the length of the holding period of the respective assets.[42] Also, if data are grouped in income brackets it is not possible to allocate the capital gains to the right individuals and in the worst case, large one-time realizations may elevate individuals with much lower incomes into a one-time high-income position distorting the true underlying distribution. Finally, the economic interpretation of the capital gain depends on what type of asset transaction it emanates from. For example, if it relates to a house sale, the sale of a closely held firm, or the execution of a work-related options program, the interpretation in terms of labor or capital income differs. Tax data typically lump together all capital gains, but in an effort to disentangle them according to the income characteristics of those realizing capital gains, Roine and Waldenström (2012) divided the top percentile incomes into work-related (earned by "working rich") and capital-related (earned by "rentiers"). They found that the "working rich" are the largest group both in terms of incomes and numbers but that its share has declined since 1980. This, however, does still not answer if realized capital gains stem from work-related activities or if high-income earners also realize capital gains in addition to their incomes.

Problems with observing and accurately dating capital gains have led many inequality researchers to exclude the realized capital gains altogether from inequality data.[43] However, in the top income literature the approach to capital gains has been pragmatic in the sense that, whenever possible, top income shares have been presented both including and excluding realized capital gains (of course making the corresponding adjustments to the reference totals). This has been possible in Canada, Finland, Spain, Sweden, and the United States. In some countries such as Australia, New Zealand, and Norway capital gains are included in the tax base but not reported separately, whereas in other countries (e.g., the United Kingdom, the Netherlands, Switzerland, and Japan), realized capital gains are not taxed under the income tax (with some variation over time) and therefore not included in the reported gross income concept.

The impact on top income shares from adding taxable realized capital gains is shown in Figure 7.6. The figure first illustrates the problem often raised with respect to including

[41] According to Haig (1921) and Simons (1938), income is the value of consumption plus any increase in real net wealth, that is, it should include all capital gains, not just the realized ones.

[42] In Sweden during 1976–1990, for gains from sales of assets held longer than 2 years only 60% was taxable, and before that all the gains from sales of assets held over 5 years were tax exempt.

[43] This is the case for the LIS, for example.

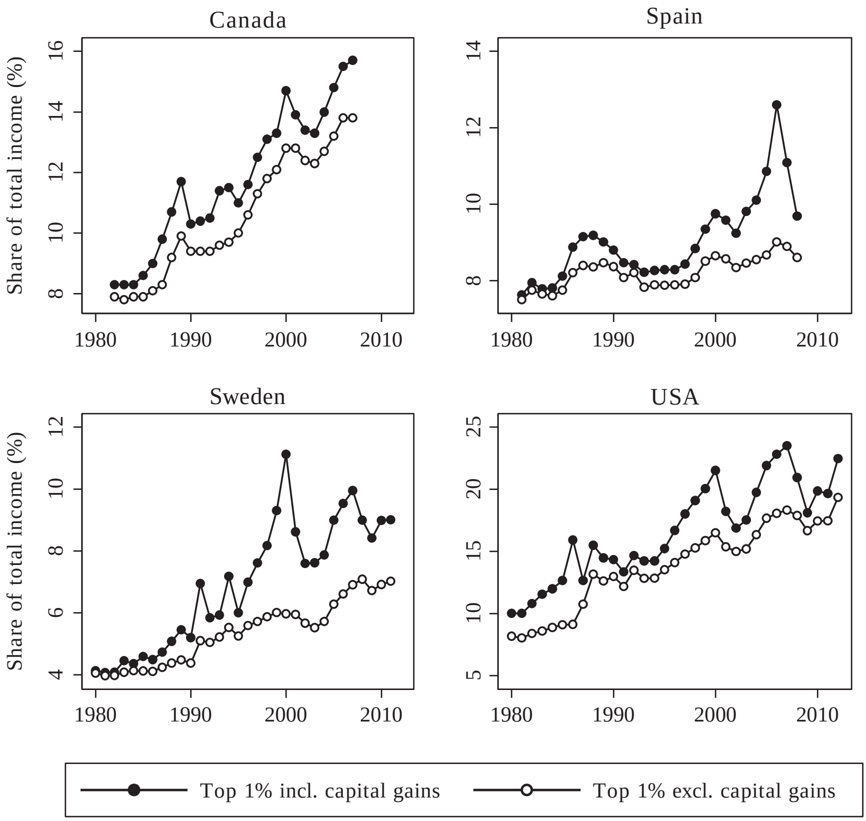


Figure 7.6 Capital gains in top income percentile, four countries. *Note: Income earners are ranked separately according to each income concept.*

realized capital gains, namely, that there are clear visible spikes in years when realizations are attractive for tax reasons. The clearest example of this is the well-known 1986 Tax Reform Act in the United States when the top percentile share was almost twice as high when realized capital gains were included, but the spikes in 1991 and 1994 in the case of Sweden are also driven by similar tax incentives.[44] But, second, even if one disregards these peak years, there seems to be a trendwise increase in the importance of realized capital gains as a source of income in the countries. Roine and Waldenström (2012) study to what extent this, in the case of Sweden, is an artifact of increasing turnover and a reflection of different individuals making occasional appearances in the top group. Using micropanel data they can compute average incomes, excluding and including capital

[44] Auerbach (1989) shows how the one-time spike was created by changed tax incentives. See Saez et al. (2012) for further references. In the case of Sweden, Björklund (1998) noted that "... due to changes in the incentives to sell stock, realized capital gains were unusually high in 1991 and 1994" and goes on to treat the values of inequality in those years as outliers.

gains, of individuals over longer time periods. Their main finding is that it is not mainly different individuals who take turns in appearing in the top group; rather it is mainly top income individuals that earn substantial amounts of capital gains in addition to their other incomes. Armour et al. (2013) and Burkhauser et al. (2013) used survey evidence from household panels in the United States and Australia, respectively, to compute both realized and unrealized capital gains and study their impact on measured income inequality. Comparing their results with those found in the top income literature for these two countries, the authors concluded that capital gains are indeed important drivers of inequality but that only using taxable realized capital gains may confuse the timing of inequality changes and also tend to overstate increases in top income shares.

Taken together, decomposing income inequality trends with respect to income source turns out to be very important for understanding the developments. Whereas earnings have always comprised the bulk of incomes of most individuals, top incomes come from both labor and capital, and variation in top income shares can largely be driven by changes in capital income flows. In the beginning of the twentieth century the highest incomes were dominated by capital income, and most of the decline is caused by decreasing capital incomes, partly due to shocks to wealth holdings during the World Wars and the Great Depression. This clearly explains some of the differences within the top that we observe in the first half of the century. In contrast, the recent upturn in top income shares is mainly due to increasing top wages and salaries, especially in the United States and the United Kingdom, but capital is also making a return in some countries.

7.2.3 The Relation Between Top Income Data and Other Measures of Inequality

As we pointed out in the introduction, the primary motivation for the top income project was a dissatisfaction with inequality data sets in general. It was a lack of comparable, annual time series of inequality over the long run that was the main problem, more than a lack of data on details within the top. As shown earlier, detailed information within the top turns out to be important in its own right and is in fact in many respects crucial for understanding the overall development. But what about the relation between top shares and other measures of inequality that cover the entire population, such as the Gini coefficient? And what about the relationship between top income shares based on tax data and similar top shares based on household surveys? This section seeks to answer these questions.

7.2.3.1 Comparing Tax-Based and Survey-Based Estimates of Top Income Shares

Household surveys are a common source for income inequality analysis. Unlike most tax data, surveys allow for household adjustments and, at times, more comprehensive income concepts. Some recent studies recalculate the U.S. top income shares of Piketty and Saez (2003) using some of the largest U.S. household surveys: the Current Population Survey

(CPS) (Burkhauser et al., 2012) and the Survey of Consumer Finances (SCF; Kennickell, 2009; Wolff and Zacharias, 2009). These studies are only able to compute estimates since the 1970s. Nonetheless, they offer valuable points of comparison for the tax-based top income share series, in particular given the potential problems with tax avoidance and other concerns related with the tax data.

The CPS-based analysis produces lower inequality levels overall and also present a lower trend increase in top shares since the 1970s. Atkinson et al. (2011), however, point out that much of this difference stems from the fact that the CPS data are top-coded, which means that the highest incomes are incompletely observed, which may underestimate the top shares. Similarly, the CPS has a lower coverage of capital gains, and given their importance in the top (as argued earlier in this chapter), this omission may account for a fair share of the difference. The survey evidence based on the SCF suffers less from top-coding and, accordingly, are more in line with the tax-based series of Piketty and Saez. In a similar comparative exercise for Australia, Burkhauser et al. (2013) contrasts the tax-based evidence with top shares calculated from the Household, Income and Labour Dynamics in Australia Survey. The authors find that top income shares are somewhat lower when using a more theoretically appropriate income concept based on the survey evidence. Regarding the overall patterns in terms of time trends and income composition, however, there is a high degree of agreement between the two sets of series. In other words, household surveys in Australia, the United States, or the United Kingdom do not seem to offer a fundamentally divergent picture from the basic evidence of the top income literature.

7.2.3.2 Theoretical and Empirical Relationship Between Top Shares and Overall Inequality Measures

To what extent can top income shares be thought of as a measure of overall income inequality? To answer this question one can refer to desirable properties of inequality measures (see, e.g., Cowell, 2011), the theoretical relationship between top shares and other inequality measures, or to the observed statistical associations between different inequality measures when based on actual observations.

As discussed by Leigh (2007), top income shares meet four basic properties that any measure of inequality should satisfy: they are not affected by any other characteristics of the population than income (anonymity), they remain the same when all incomes are multiplied by the same number (scale independence), top shares remain unchanged if the population is replicated identically (population principle). When it comes to the transfer principle, this is only satisfied in its weak form because a transfer from a high-income individual to a low income never increases the measure, but it may remain unchanged. A transfer from the top group to the rest of the population lowers the top income group share, but transfers within the respective groups leave the measure unchanged. A direct consequence is, of course, that top income shares cannot capture changes that happen *within* the lower part of the distribution.

What is the quantitative impact of a top income share change on the Gini coefficient? Atkinson (2007) suggests a useful approximation. If we assume that the top share is negligible in size but has an income share S, the total Gini coefficient (G) can be approximated as $G = S + (1 - S)G'$, where G' is the Gini coefficient for the population excluding the top group. To use the example given by Atkinson (2007), if the Gini in the rest of the population remains at 0.4 but the top percentile group experiences a 14 percentage point increase in their share (as in the United States between 1976 and 2006) this leads to an 8.4 percentage point increase in the overall Gini.

What about the correlation between top income shares and Gini coefficients in data? Figure 7.7 illustrates the overall, average relationship the 2 for 16 developed countries. The left panel illustrates a positive and high correlation, 0.86, between the *levels* of inequality. The right panel shows that the correlation between average annual inequality *changes* during the period 1985–2005 is lower but still positive and high, 0.57.

Looking at the relationship more systematically, Table 7.4 gives a correlation matrix for the relation between top income shares and broader measures of income inequality. Using data from the LIS, the WIID, and the WTID over the past 30 years, the table shows Pearson correlations between three top income shares (the top percentile, the top decile,

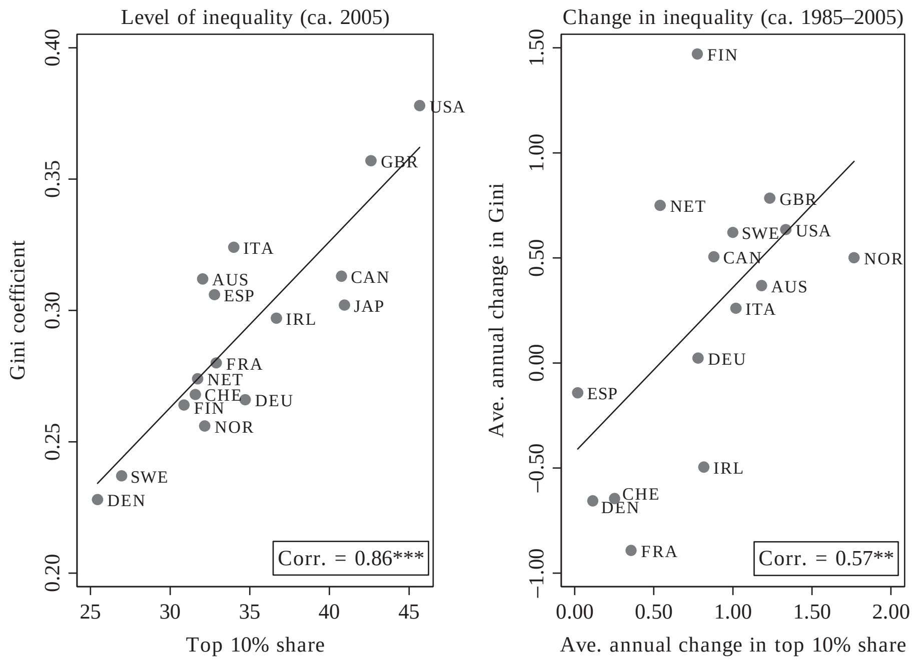


Figure 7.7 Top income decile and the Gini coefficient. *Source: Gini coefficient for disposable incomes of equivalized households are retrieved from the Luxembourg Income Study Datacenter (www.lisdatacenter.org) and top decile gross income shares from the World Top Incomes Database. Pearson correlations are statistically significant at the 1% (***) and 5% levels (**), respectively.*

Table 7.4 Correlations between top income shares and other inequality measures

	Top 1% (P99–100)	Top 10–1% (P90–99)	Top 10% (P90–100)
World Income Inequality Database (WIID)			
Gini coefficient	0.50	0.25	0.42
Luxembourg Income Study (LIS)			
Gini coefficient	0.62	0.69	0.73
Atkinson index ($\varepsilon = 0.5$)	0.61	0.65	0.70
Atkinson index ($\varepsilon = 1$)	0.53	0.61	0.64
P90/10	0.59	0.70	0.72
P90/50	0.57	0.65	0.68

Notes: The correlations are all statistically significant at the 1% level. The number of observations for the WIID variables is 300 for Top 1% and 263 for the Top 10–1% and Top 10%, and 63 for all LIS variables.

and the lower nine percentiles in the top decile) and the Gini coefficient, the Atkinson index using two inequality aversion parameters, and the income ratios between the 90th percentile and 10th percentile (P90/P10) and the median (P90/P50). The correlations are the lowest for the WIID Gini coefficients, 0.25 and 0.42 for two of the top share measures. When using the LIS data, correlations are markedly higher, between 0.53 and 0.57 for the top percentile and between 0.64 and 0.74 for the two other income shares.[45]

Finally, we also examine what the relationship between top income shares and the Gini coefficient looks like over the very long run. We do this by plotting series for two countries, the United Kingdom and the United States, where the Gini coefficient spans the entire period since the beginning of industrialization until present day, whereas the top income percentile only covers the last century. Figure 7.8 shows the results from this exercise. The evidence suggests that the twentieth-century experiences are quite similar across the two indices of inequality. In both countries the documented equalization appears in both measures with only minor deviations in the magnitudes. These observations thus indicate that had we accessed top income data for the eighteenth and nineteenth centuries they may have generated similar long-run trends since the 1700s as those portrayed in Figure 7.8, but of course we cannot make any certain statements without hard evidence.[46]

Altogether, this section shows that top income shares are related to well-known measures of overall income inequality such as the Gini coefficient, the Atkinson index or

[45] We also find strong "conditional correlations" using regression analysis where we account for time trends and country effects, similar to what is done by Leigh (2007) and by Smeeding et al. (2014) in Chapter 8 of this Handbook.

[46] In fact, both Lindert (2000) and van Zanden (1998a,b) seem to find at least some cases of a deviation in inequality trends in the United Kingdom and the Netherlands, respectively, across elite status and population-wide measures. See also the study of U.S. income inequality trends of Lindert and Williamson (2014) for similar evidence.

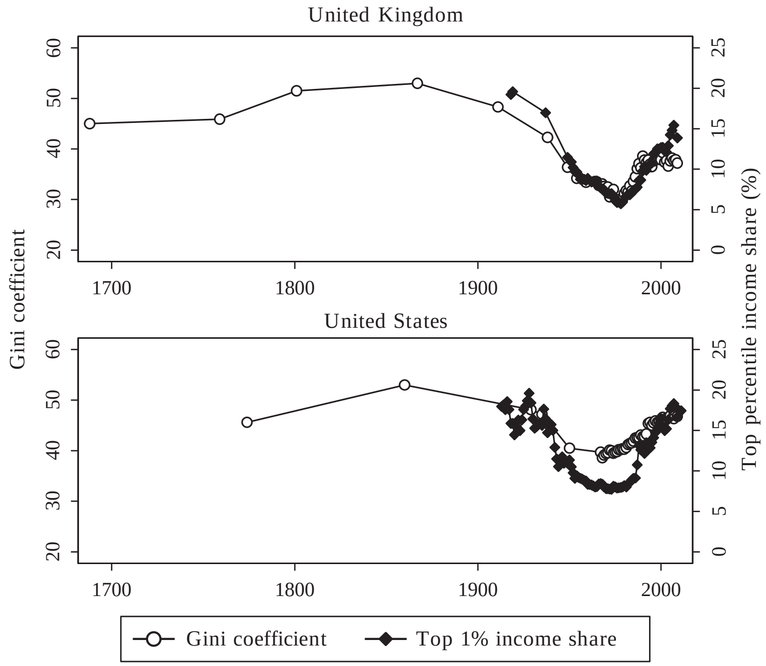


Figure 7.8 Long-run inequality trends using Gini and top income percentile share. *Sources: Gini coefficients for the United Kingdom from Lindert (2000, table 1), Milanovic (2013, table 1), and Office for National Statistics (2011, table 5), and for the United States from the same Lindert and Milanovic sources and U.S. Census Bureau (2011, table A-3). Top income shares from World Top Income Database.*

income ratios, both theoretically and empirically. Top income shares fulfill properties for being sensible inequality measures and quantitatively changes in top shares have a nontrivial impact on the Gini coefficient. They are also significantly correlated with overall measures of inequality although they (by definition) do not capture variation within the lower part of the income distribution. Does this imply that we can uncritically assume that top income shares can serve as a proxy for, say, the Gini coefficient? No, of course it does not. The correlations we present rely on evidence from time periods when we observe both top shares and enough data to calculate the other inequality measures. In practice, this means relying on data starting in the 1970s. In the few cases when we have data for longer periods these confirm the close relationship when going back in time. However, as shown by Smeeding et al. (2014) in Chapter 8 of this Handbook, the relationship is weaker in recent decades as household surveys do not fully capture the developments in the very top of the distribution. In the end, how to use top shares (or any other summary statistic) when aiming to capture overall income inequality, is a question of judgment. Our view is that, based on the evidence we have, and, in particular, given the restrictions in terms of available alternatives, top shares should not be dismissed as being "only about the top" but are also useful as a general measure of inequality in over time.

7.2.3.3 Other Series over Long-Run Inequality: Wages, Factor Prices, and Life Prospects

Much of what we write in this chapter is based on the assertion that the long-run evolution of income inequality is meaningfully reflected in the evolution of top income shares, that is, the shares of income accruing to top fractiles in repeated annual cross-sectional income distributions. Notwithstanding our conclusions in the previous section, there are some important limitations to the top income data, and it is therefore useful to complement these series with alternative measures. One is the poor coverage of period before 1900; top income data only exist in a handful of countries, none earlier than the 1860s and in most cases only in the form of a few scattered year observations. Furthermore, top income data are not ideal to study the dynamics between inequality and economic development in relation to industrialization as characterized by some theories such as the Kuznets hypothesis. Last, the use of repeated annual income distributions prevents conclusions about trends in the distribution of lifetime incomes, that is, whether differences in the quality and length of people's life span has changed in such way that the overall inequality trends are either mitigated or boosted depending whether it is the lives of the poor that has improved the most or the least.

In this section, we present some additional evidence on long-run inequality that have bearing on these issues. We do this by studying trends in some other measures that are popular in the past literature: wage dispersion across occupations (and regions), factor price differentials, and differences in life prospects.

The first measure, wage dispersion, is most often constructed as the wage ratio of rural to urban workers or of professionals (skilled) to blue-collar (unskilled) workers. Besides being available over very long time periods, often well before industrialization, these measures also offer a closer association with the original Kuznets conjecture, which was about changes in wage inequality precisely between urban and rural workers within countries over the path of industrialization. A large number of studies have scrutinized this conjecture using different types of wage ratios, and they offer somewhat contradictory evidence (also see Section 7.4.1). In his review of this extensive literature, Lindert (2000) asserts that, at least concerning the United Kingdom and the United States, historical series are still too incomplete to allow for any firm conclusions. However, at least they do not establish any clear support for strong increasing trends in sectoral or occupational wage differentials as Kuznets' assertion would stipulate.[47] In a study of the evolution of skill premia across occupations during the premodern era up until the early

[47] There are numerous measurement problems that researchers have dealt with. These include how to deal with nonmonetary reimbursement that was particularly common in agricultural professions, or the living conditions and health risks exposing workers differently in cities and on the country side. Specifically, Lindert (2000) points out that whenever costs of living differ between rich and poor, the dispersion of real wages differ from the dispersion of nominal wages. Lindert points to evidence from eighteenth-century England that the cost of living fell slower for the lowest 80% than for the top 20%, indicating that real inequality increased more than nominal inequality.

twentieth century in the entire Western world, van Zanden (2009, ch. 5) also failed to find any evidence of increased wage dispersion during industrialization. Looking instead at the twentieth century, wage ratios decline almost unanimously in Western countries. Not only does this development fit the acclaimed downturn of the Kuznets curve but it also correlates positively with the inequality trends suggested by the declining top income shares. As Lindert (2000) emphasized, however, the twentieth-century drop in pay differentials does not seem to be driven by the forces suggested by Kuznets. Instead the factors compressing wage ratios were rather aligned to institutional developments such as labor market regulations and the expansion of trade unions and to the extension of educational attainment for large masses in the population (Goldin and Katz, 2008).

Sweden has in the past literature been referred to as a "clear example of the Kuznets curve" (Morrison, 2000, p. 227), an assertion based largely on Söderberg's (1991) investigation of sectoral wage dispersion. Swedish wage differentials across skilled and unskilled workers seem to have risen between 1870 and 1930, with exception for a sharp drop during the First World War, and then turned downward until 1950. As Sweden's industrialization can be said to have begun around 1870 and peaked around the turn of the century, the skill differential in wage indeed matches the Kuznets pattern. However, more recent research using new evidence on wage differentials between rural and urban workers (Bohlin et al., 2011) and across occupations (Ljungberg, 2006) cannot replicate these results. They find either no trend at all or even a negative trend beginning already in the nineteenth century, casting doubts about the existence of even a Swedish Kuznets curve.[48]

Relative factor prices, typically expressed as the ratio of land rents to real wages, represent another outcome that bears information about inequality trends, even if it is primarily motivated by trade theory. One basis, the inequality interpretation, is offered by Lindert (1986, 2000), who argued that land ownership during the nineteenth century was highly concentrated and changes in its return relative to real wages can reflect changes in the overall income inequality. According to several studies (Clark, 2008, p. 274; Lindert, 2000; O'Rourke and Williamson, 1999; van Zanden, 2009), the wage–land rental ratio did not decrease (i.e., inequality did not increase) at all during the nineteenth century in the industrializing world. If anything, the wage–rental ratio went up in the decades before the First World War, but whether that reflected a true equalization

[48] Specifically, Bohlin et al. (2011) compared the wage gap between agricultural (rural) workers and engineering (urban) workers between 1860 and 1945, controlling for differences in nonwage reimbursement and costs of living. They find no secular trend in the wage gap before 1950 but a considerable short-term responsiveness to shocks, for example, to local living costs. Ljungberg (2006) compared wages of male manufacturing workers with wages of graduate engineers, college engineers, and secondary school teachers between 1870 and 2000, finding that unadjusted wage gaps trended downward but that the pre-First World War trend largely disappeared when controlling for the growth of human capital in the labor force at large.

in the midst of the second industrial revolution or merely the demise of the rural land owners remains an open question.[49]

Finally, although the dispersion of incomes earned during a single year is often a relevant time frame of analysis, there are dimensions of personal welfare when outcomes over longer time spans are of primary concern. If, for example, industrialization allowed the broad masses to live better, eat healthier, and work safer, and thereby live longer, without affecting the lives of the rich at all, this would result in an equalization of lifetime incomes even if distribution of annual incomes did not change at all. The literature on differential mortality trends over the long run and their implications for lifetime income inequality trends is quite small. In his review article, Lindert (2000) referred to studies of the United Kingdom that seem to reach conflicting conclusions, some finding that the biggest gains in life expectancy materialized among the already rich, whereas others found the opposite. Clark (2008) looked at the differences in life prospects between "rich" and "poor" before and after industrialization, broadly put. He found that the rich–poor difference in terms of male stature decreased from 3% to 1%, in life expectancy from 18% to 9%, in number of surviving children from 99% to −19%, and in literacy from 183% to 14%.[50] However, the most recent research on socioeconomic inequalities in death over the long run presents a more sceptical view of the role of industrialization. Using historical longitudinal microdata from several countries aiming at uncovering the causal impact of industrialization on social mortality differences, scholars have not found any clear trend break along with the industrialization and, in general, no clear impact of income on mortality at all.[51]

Altogether, the evidence put forward in this subsection has broadened the focus on long-term trends to also include other measures of inequality such as occupational wage ratios, factor price differentials, and lifetime-amended income inequality. These other distributional sources offer insights into pre-1900 inequality trends, the economic dynamics more closely related to the Kuznets conjecture, and into the development of the inequality of lifelong well-being, all of which are unsatisfactorily addressed by the top income data (and not addressed at all by other pre-1900 income inequality data sources). The main message from these studies is that there are few indications of an increase in inequality during the nineteenth century, that is, the era when most Western countries experienced their definitive industrial takeoffs. There is hence little empirical support for the first part of the Kuznets inverse-U curve. We would still hesitate to extrapolate our top income shares backward into the nineteenth century based on the

[49] O'Rourke et al. (1996) established the overall trends in the wage–rental ratios, arguing for a crucial role of trade openness as driver of the equalization, whereas Clark (2008, p. 274) emphasized the fact that land owners failed to keep up with productivity booms in the industrial sector.

[50] See Clark (2008, table 14.4, p. 283), based on a number of different sources.

[51] See Bengtsson and van Poppel (2011) and the references therein.

evidence from pay ratios. In terms of lifetime income inequality movements, there is again no clear trend that deviates notably from the one offered by the top income shares. If anything, the twentieth-century equalization may be even stronger if would adjust for changes in longevity differences across the distribution, but this conclusion rests on still quite tentative evidence.

7.2.4 Income Inequality over the Long Run—Taking Stock of What We Know

Combining all the preceding information, it seems that there are three possible permutations of broad overall trends since the beginning of Western industrialization. To continue the letter-analogue to describe shapes, the question is if we (with a bit of imagination) see an N, a U, or an L. The N-shape corresponds to an increase in inequality over industrialization followed by a decrease over the twentieth century and again an increase since around 1980. The U-shape would be a situation where inequality is high before and during the period of industrialization, then declines over the twentieth century, and increases again after around 1980. Finally the L-shape corresponds to the U-shape but without the upturn around 1980.

The question marks, thus, revolve around to what extent there was an increase or not during industrialization and to what extent there has been an increase in recent decades. The answer to the first question is difficult due to lack of clear evidence. There are some signs of increased inequality during industrialization but many studies also point toward high and relatively stable levels before the decrease in the twentieth century.

When it comes to the second question about the increase since around 1980, the evidence is much more solid and clearly indicates that the answer depends on the country in question.[52] In some countries, especially the United States and the United Kingdom, inequality has risen sharply. This increase has taken place from a level that was already high in relation to others before it started. In countries like Sweden and Finland, increases have also been substantial but here from internationally low levels to levels that are much higher, but remain among the lowest. In other words, the increase in percentage terms has been almost as large in Sweden and Finland as in the United States and United Kingdom but the level difference is very significant. In some other countries, for example France, Germany, and Japan, there is no clear upward trend but in absolute terms inequality remains higher than in the Nordic countries.

To what extent is this picture any different than the one we had before the top income literature and other findings that emerged in the past decade? In terms of the broad overall developments, it may actually not be so different. There are some more studies suggesting that the increase in inequality during industrialization is not so clear

[52] For a closer analysis of the post-1970 inequality trends in the industrialized countries, see further Chapter 8 in this Handbook (Smeeding et al., 2014).

and the recent upward trend in inequality has been made even clearer. Of course we have a lot more data on inequality over the long run in the form of top income shares. But overall there is nothing dramatically new in terms of the secular trends over the long run.

What is new, however, is the change in our *understanding* of these trends as a result of a number of features in the top income data. First, the detailed analysis of changes *within* the top of the distribution has shown just how much of the development is driven by the top 1% group of the distribution and, conversely, how surprisingly stable the income share of the lower half of the top decile has been over the long run. Second, the decomposition of income according to source has increased our understanding of the importance of accounting for all sources and how the same broad trend could be driven by entirely different mechanisms depending of the development of capital and wages, respectively. This applies both to the aggregate economy and to different groups across the income distribution. Third, the often yearly observations have shown the importance of sufficiently high frequency data. In particular, this aspect of the new series has been an important part of the focus on the role of shocks and war especially in the first half of the twentieth century, thus creating an at least partly new interpretation of the decline in inequality in the first half of the twentieth century. Finally, the relationship between top incomes and other measures of inequality illustrate how this literature has contributed *both* to our understanding of the importance of developments within the top, and the possibilities to use these measures as proxies for overall inequality.

7.3. LONG-RUN TRENDS IN WEALTH INEQUALITY

It is fair to say that the majority of research on economic inequality has focused on incomes. Much less attention has been given to the role of wealth, which is unfortunate for a number of reasons. As a determinant of people's consumption possibilities, personal wealth is of first-order relevance. The classical Haig–Simons definition of income states that income is what we can consume while keeping our real wealth intact. Wealth can also determine which opportunities individuals have to make investments and pursue different occupations, especially in the presence of credit constraints. The interplay between the distribution of wealth and development is also central to many theories attempting to explain the cross-country differences in long term development.

This section presents and discusses the existing research on the long-run evolution of wealth inequality. The ambition is to harmonize the outline with the previous section on the long-run trends in income inequality. We begin by presenting the core methods and data issues concerning how to measure wealth, wealth inequality, and how to tackle the specific challenges associated with studying historical trends. Thereafter we present ten country case studies for which we have sufficiently good data on wealth concentration for at least a century and in some cases from the beginning of each country's industrialization.

Finally, we bring together the pieces of evidence into cross-country mappings of the trends, searching for common patterns that may help us address the overall questions about the relationship between economic development and inequality.

7.3.1 Data and Measurement

Despite the arguments for studying wealth and its distribution, the empirical literature on wealth inequality is still limited, particularly when it comes to the long-run perspective. Naturally, there are many reasons for this past neglect, but the problem of agreeing on a manageable definition of wealth and then the practical problems associated with measuring it empirically are most likely important.

Sources for studying wealth over time are of different sorts. In their investigation of the analysis of wealth distribution, Davies and Shorrocks (2000) pointed at the five most common sources of wealth data: wealth tax returns, estate tax returns (or probate records), investment income method (using capital income and some assumed or observed net rate of return), household surveys, and journalistic rich list. With respect to investigations of long-run patterns, perhaps the most consistent of these sources is estate records. They have existed for centuries with largely the same basic structure of assets and debts of the deceased individuals. Unfortunately, there are few compilations of estate records in most countries over time, which is why we still lack data on wealth distribution from this source. A few countries have presented tabulated sizes of estate records in relation to estate tax compilations. Wealth tax statistics is another common source, available in a fairly homogenous way in several countries over long periods of time. Here, however, the problems of what components are included in the tax base or how large share of the population that is covered in the statistics are more pressing problems. Surveys, finally, comprise a more recent source for wealth distribution evidence.

Historical evidence on wealth distribution data is primarily based on wealth and estate taxation statistics. These fiscal instruments have been used for centuries and offer consistent source materials. Authorities have often also been interested not only in collecting the revenues but also in calculating the size of each tax base as well as their respective size distribution. Of the historical evidence presented later, series from France, the United Kingdom, and, in part, the United States, all emanate from the estate tax and, specifically, samples of individual estate tax returns. U.S. wealth distribution data from the latter part of the twentieth century and the beginning of the twenty-first century are also available in household surveys. Wealth distribution data from Denmark, the Netherlands, Norway, and Switzerland are all based on wealth tax statistics, in most cases as tabulated distributions published by each country's tax authorities. For Finland and Sweden the bulk of the data come from both wealth tax statistics but there are some complementary observations from estate tax returns. For Australia, finally, observations come from estate tax data, wealth surveys, and even journalistic rich lists.

7.3.1.1 The Wealth Holding Unit

The concept of wealth owner varies across the empirical studies covered in this chapter depending on the nature of the data source used. When wealth tax-based data are used (as in the Netherlands, Switzerland, and the Nordic countries), the most common unit of observation is *households*. For the most part, this means tax households where married couples (and their under-aged children) count as one, as do children 18 years or older living at home. Many of the survey-based wealth records from recent decades, however, define households as cost-based households, the major difference being that adult children living at home are now included in the parents' household. When studying very long time spans, households sometimes also included servants, parents or grandparents, slaves, or unregistered immigrants. Shammas (1993) shows that the U.S. historical wealth concentration is sensitive for the treatment of these different subgroups into the reference tax population.[53] Estate tax data and probate inventories (used in France, the United Kingdom, and the United States) are instead based on (deceased) *individuals*.[54] Most studies focus on adult individuals, thereby imposing a lower age cutoff normally between 15 and 25 years of age.[55]

To define wealth holding units consistently matters for the distributional estimates. As was pointed out earlier in the discussion of the distribution of incomes, individual-based data tend to (but must not) give rise to a more unequal wealth distribution than does the household-based data (Atkinson, 2007). Roine and Waldenström (2009) compared shifts in Swedish top wealth shares using household and individual distributions finding no important differences, and Kopczuk and Saez (2004) reached the same conclusions in their analysis of U.S. wealth distribution trends.

7.3.1.2 The Concept of Wealth

The definition of personal wealth that is most commonly used in studies of wealth distribution is *net wealth*, also called net worth or net marketable wealth. Net wealth consists of the sum of all nonhuman real and financial assets less debt. Real (or nonfinancial) assets primarily consist of housing and land, but they may also include durable consumption goods (see further the discussion later), for example, cars, boats, furniture, and also valuables such as antiquities, jewelry, and art. In the distant past, even items such as clothing and other semidurable consumption goods were often inherited (especially among the less wealthy) and may also be covered among the nonfinancial assets. Financial assets are cash, bank deposits, corporate stocks, bonds and other claims, and insurance savings,

[53] Shammas shows that when including slaves in the population of wealth holders, the top percentile wealth share increased by 15% (Shammas, 1993, table 1).

[54] Some estate reports include joint property if there is a surviving spouse and the property of a deceased spouse that has not previously been transferred to heirs.

[55] Variation in age cutoff across countries and even within countries over time may introduce problems of comparability (Atkinson and Harrison, Chapter 6).

which today also include some parts of funded pension assets. Debts, finally, are the sum of housing mortgages and loans for consumption, investment, or education.

As already stated, our definition of wealth does not include peoples' inherent or acquired skills, or human capital. This is a natural implication of the wealth definition set out at the beginning, which focuses on assets that are *marketable* and thus possible to sell or purchase at a market place.[56] Historically, such market for human wealth has existed, namely in association with slavery. In terms of aggregate wealth the total value of "slave assets" was somewhere between 15% and 30% of total national wealth (Piketty and Zucman, 2014, figure 11; Soltow, 1989, p. 180). According to Soltow (1989, p. 267), slaves were disproportionately held by the wealthy, and the inequality in slave ownership was almost three times as large as the inequality in land and dwellings.

Measuring net wealth is sensitive to the valuation of assets. Ideally assets should be valued at current market prices, net of taxes and transaction costs, the theoretical reason being the possibility to convert wealth to consumption. However, most estimates of historical inequality use data where assets are reported in tax-assessed values rather than in market values. Tax laws are typically designed to strike a balance between the revenue needs of government and tax collectability of tax authorities, and the rules regarding asset coverage or valuation criteria may thus not be aligned with what researchers would ideally like to have. But if the discrepancy across tax and market values is similar across the distribution—and historically we think that this was arguably often the case—the biasing effect of valuation on relative wealth *shares* should be small. Only a few studies have delved into these questions. Examples are the analyses of inequality trends in the United States where Williamson and Lindert (1980a,b) and Wolff and Marley (1989) investigated whether tax-driven avoidance distorts the use of tax data for distributional analysis (and they generally found that it does not). Atkinson and Harrison (1978) examined how the valuation of taxed assets may influence inequality, for example, looking at life policies (table 4.6) and offshore assets (pp. 161f).[57] Roine and Waldenström (2009) studied the effect of valuation by using several alternative estimates of aggregate wealth (based on either tax or market values as well as including items that have not been taxable) and also different assumptions about the distribution of the difference between these alternative reference totals and the baseline specification. They found that there are some differences in the levels of wealth shares over the period, but that the trends in wealth concentration remain unchanged. Altogether, we believe that the comparability of the estimated shares presented in this chapter is good over time.

[56] Some scholars have still tried to quantify the value of individuals' lifetime human capital and its distributional characteristics. Reviewing these estimates and their trends over time, Williamson and Lindert (1980b, p. 71) came to the conclusion that "(w)e have, then, two reasons for believing that trends in conventional wealth distributions understate the true leveling in total wealth distributions." Whether this result is stable across countries and over longer time periods is an open issue worthy of further inquiry.

[57] See also our later subsection on the role of tax avoidance and evasion.

Some components are especially difficult in the analysis of personal wealth. Although some of them appear in the wealth data in several countries and time periods, their presence is associated with uncertainty concerning both valuation and conceptual adequacy. In the following we discuss three of the most important "problematic assets" and how they are typically treated in the historical sources. In the end, they do not, however, affect the main conclusions about the long-run inequality trends reported later.

(i) Pension and Social Security wealth is a composite term for the net present value of individuals' entitlements to future private and public payments for pensions and other social outlays. These assets are for the most part not included in the historical inequality estimates. Conceptually, scholars have shown that expectations about future public pensions reduce the incentive to accumulate private wealth (see, e.g., Berg, 1983; Feldstein, 1976; Gale, 1998), and thus a comparison of private wealth across systems with differing public pension coverage may be misleading unless retirement wealth is accounted for. Researchers therefore sometimes add Social Security wealth to the net marketable wealth of households, yielding a concept often called *augmented wealth*. Studies of the concentration of augmented wealth typically find that it is substantially lower than the concentration of marketable wealth. For example, Wolff (2007) found that the Gini coefficient for the United States in 2001 dropped by a fifth when going from net worth to augmented wealth, and Frick and Grabka (2013) found a similar drop for Germany in 2007. The Inland Revenue in the United Kingdom presented for many years series of the distribution of marketable wealth (Series C) as well as wealth including public and private pension entitlements (Series E), exhibiting Gini coefficients that were about a third lower when including pensions.[58]

However, there are numerous problems associated with defining pension assets, or other "drawing rights" on the Social Security system, as private property, and until questions like those are fully settled we will not see a comprehensive treatment of pension and Social Security wealth alongside net marketable real and financial assets. The main issue is how to judge the fact that, on the one hand, not having the public system would have required an individual to save privately, thus decreasing consumption possibilities, but on the other hand, the "drawing rights" are not marketable wealth and cannot be converted freely into other consumption by the individual.[59]

[58] See Inland Revenue, *Inland Revenue Statistics 1985*, London, HMSO, table 4.8.

[59] For example, pension assets are not fully accessible to their owners on demand at any time (they are not possible to realize before retirement). Furthermore, they are partly defined in collective forms and are hence not well defined for all individuals (or households) even within the system. The calculation of current claims on future pensions necessitates a number of complex assumptions about people's life expectancy, future rates of return on the capital markets, and so forth. There is also a mix of public and private pensions, some being funded and others unfunded. Finally, it is not obvious where to draw the line in terms of valuing the rights of citizens' claims on the public sphere: how valuable is the claim on childcare, elderly care, unemployment insurance, and even the right to freely travel on public roads or be protected by the country's military defense.

(ii) Consumer durables are not always included in the wealth data, and when they are their valuation is difficult. First of all, this asset class is typically completely absent from wealth tax returns or administrative tax registers, primarily for evasion reasons. It is thus not part of the bulk of the distributional estimates examined in this chapter. However, insofar as data are based on probates or estate tax returns or household surveys, durables are more likely to be included because of smaller possibilities (and smaller incentives) to evade.[60] Atkinson and Harrison (1978, p. 43) noted that the valuation of consumer goods is difficult, and they often take too low values in estate data. In general, it is actually an open question whether consumer durables should at all be included in the household balance sheet. According to the System of National Account they should not because all consumed goods are assumed to depreciate within 1 year and therefore cannot contribute to any fixed asset formation.[61] However, many durables (e.g., cars, boats, and some electronic equipment) arguably last more than 1 year, and for this reason some countries (such as the United States) do include durable consumer goods in household balance sheets. Historically, consumption goods like china, furniture, and even clothing were important parts of household inventories and were inherited along with other assets. Waldenström (2014) estimated the household balance sheet of Swedish households since 1810, finding that durables represented between 10% and 20% of nonfinancial assets throughout the period up until today. Interestingly, durables grew more important in the middle of the twentieth century, which is related to the growth in earnings potential of increasingly educated middle-class households (Roine and Waldenström, 2009).

(iii) Foreign wealth holdings have historically been sizeable in many countries, especially colonial powers such as France and the United Kingdom. In a recent investigation, Piketty and Zucman (2014, table A27) found that net foreign wealth represented between a 10th and a quarter of total national wealth in these two countries from the middle of the nineteenth century up to World War I. At the individual level, information about foreign assets such as foreign government stock and bonds and other real estate is most likely completely absent from domestic wealth tax returns, but should in principle be more visible in estate data. As noted by Atkinson and Harrison (1978, p. 161), however, overseas real estate was not taxable before 1962 and therefore not included in wealth inequality estimations before this year. In an attempt to gauge the importance of the acclaimed tax-driven capital flight from Sweden during the period from the 1970s to the 2000s, Roine and Waldenström (2009) used residual flows in the Balance of

[60] An example of this was shown by the Swedish public investigation Kapitalskatteberedningen (SOU, 1969, table 78, p. 276) in a sample of estate tax returns matched with the deceased individuals' last wealth tax returns (from the preceding year). "Other assets" (*Övriga tillgångar*) were four times larger on the estate tax returns, and their largest component, "inner inventories" (*Inre inventarier för personligt bruk*; durable consumer goods, art, antiquities, etc.), was missing altogether from wealth tax returns.

[61] In the realm of corporations, consumption is viewed as firms' running expenses instead of as investments; only the latter results in an accumulation that forms a stock of assets.

Payments and Financial Accounts to estimate the aggregate offshore wealth held by residents. Assuming that this wealth was primarily held by the richest residents, the authors found that the top percentile wealth share rose from about 20% in the 2000s to over 30%, depending on assumptions about interest rates on foreign capital and whether to include the closely held corporations of superrich Swedes.[62] Also, without explicit reference to distributional aspects, Lane and Milesi-Feretti (2001, 2007) constructed estimates of the external wealth of nations since 1970. However, going further back we know very little about the role of offshore wealth in historical eras and can therefore not offer a consistent interpretation of their role for long-run inequality trends.

7.3.1.3 Measuring Historical Wealth Inequality

When we estimate the concentration of wealth, we use a similar methodology as when calculating top income shares. That is, we estimate the *wealth share held by various fractions of the population* by dividing the observed top wealth holdings for specific groups (fractiles) in the top by a reference total for all personal wealth in the economy. Just as in the case of historical income distribution data, the historical wealth distribution data often come in the form of tabulated distributions of grouped data. This means that we observe wealth holders and their net wealth divided into different wealth size classes. To get the exact wealth share accruing to certain fractiles in the top, such as the top percentile or the top decile, we use the Pareto interpolation technique described previously.

Using top wealth shares as measure of inequality has several advantages for our purposes.[63] Most historical sources of wealth data come from wealth and estate tax returns, and the group most consistently represented in these tax listings throughout history is the rich (i.e., where the wealth was), which makes them the most homogenously observed group over time. Moreover, wealth distributions are heavily skewed—much more so than income distributions—and top wealth holders have often held the vast majority of all personal wealth; between 70% and 90% before the Second World War and between 50% and 70% thereafter. Studying the top and its wealth therefore means that almost all personal wealth is being studied. Finally, most of the historical wealth inequality estimates constructed by past researchers come in the form of top wealth shares, especially as top wealth percentiles, and this measure is therefore the most appropriate to use for our purposes.

[62] An alternative approach to estimate the hidden wealth of nations was proposed by Zucman (2013), who instead used balance sheet statements of countries' portfolio investments in the 2000s to detect systematic mismatches that could be interpreted as evaded capital, presumably of the rich. Zucman did, however, not present any distributional implications of his calculations.

[63] Having said this, other measures of wealth inequality may be applicable on the available long-run evidence. In particular, a headcount measure based on the number of wealthy over some cutoff line, which could be defined as a multiple of average incomes, has been proposed by Atkinson (2008b).

A specific challenge associated with estimating top wealth shares is the measuring of the reference total of net wealth of the whole population. Wealth tax data typically only cover the top households that have paid wealth tax, and researchers must therefore limit their observations to years when attempts to measure the corresponding total for the whole population have been made, for example, in censuses or special public investigations. In the case of Sweden, for example, there are years for which tabulated top wealth data exists, but there is no reliable information about the reference total wealth to be found. Estate data also have problems with constructing population measures but they are of a slightly different kind. Researchers here typically try to collect a sample of estates that is representative for the whole population, which thereby enables them to compute the relevant inequality measures using only the sample at hand. However, most of the time the estate data sources are themselves not fully representative for the population, mostly lacking information about people with low levels of personal wealth.[64]

The different wealth data sources also display the wealth distribution for different entities. Whereas wealth tax data and surveys reflect the distribution of the living population, estate tax data and probate inventories reflect the distribution of the deceased. Because those people who die during a year are not a representative sample of the living population (e.g., because the old are heavily overrepresented), these two distributions are not immediately comparable. The usual procedure used by researchers to make them comparable is by applying so-called mortality multipliers, which are inverse mortality rates for different age, sex, or social status groups.[65] In this way, the distribution of estates can be transformed so as to reflect the wealth distribution among the living population.

7.3.1.4 Tax Avoidance and Evasion

As already noted, using data from administrative tax-based statistics to compute measures of wealth distribution gives rise to some problems relating to tax evasion and avoidance. But, as in the case of the income distribution, the extent to which such activities lead to errors in estimated wealth shares is, however, not clear. If noncompliance and tax planning is equally prevalent in all parts of the distribution—it may, of course, take very different forms—this affects the reported wealth levels but not the shares. The same goes for comparisons over time and across countries (see Section 7.2.1.5 for more on this). Unfortunately there is little systematic evidence on this. There are overviews, mainly

[64] The estate tax returns used to calculate U.S. wealth shares over the twentieth century only cover the two richest percentiles in the entire population, and reference total wealth was collected from national balance sheets (Kopczuk and Saez, 2004).

[65] The methodology of using inverse mortality rates, preferably adjusted for sex and social class, was pioneered by the works of Coghlan (1906) and Mallet (1908) and was also implemented around the same time by the Swedish statistician Isidor Flodström (Finansdepartementet, 1910). For a detailed account of the mortality multiplier methodology and theoretical underpinnings, see Atkinson and Harrison (1978, chapter 3).

concerned with personal income taxes, suggesting that, although avoidance and evasion activities are important in size, there are no clear results on the incidence of overall opportunities or on these activities becoming more or less important over time.[66]

Moreover, it is not clear whether to expect more or less avoidance and evasion in countries with higher tax rates. As the incentives to engage in avoidance and evasion become higher when taxes increase, so do the incentives for tax authorities to improve their control.[67] Regarding wealth and estate taxes it may seem plausible to think that estate tax data are more reliable because it is typically in the interest of the heirs to formally establish correct valuations of the estate.[68] At the same time tax planning aimed at avoiding the estate tax is an important industry in the United States and elsewhere. This may affect the reliability of the data. For wealth, tax data problems of underreporting are likely to be similar to those for income data, with items that are double reported being well captured, whereas other items are more difficult.

Finally, the use of tax havens may be a problem, and as we discussed earlier there are indications that substantial amounts have been hidden over the past decades (see for example, Johannessen and Zucman, 2014, and references therein). Given the large fixed costs related to advanced tax planning, it is likely that such activities are limited to the very top of the distribution. If this has become more important over the past decades—something that seems likely—then estimates of wealth concentration for recent periods may understate wealth holdings in the very top and not be directly comparable with estimates produced for earlier years in this century, in particular top wealth shares may be underestimated for recent decades.

7.3.2 Evidence on Long-Run Trends in Wealth Inequality

In this section we present evidence on the evolution of wealth inequality in 10 Western countries. The length and detail of the series vary but in most cases the first observations are from around 1800 and with relatively frequent observations throughout the whole of the twentieth century. The relatively small number of countries for which we have data allows us to delve a little deeper into each country case, examining the specificities associated with national histories as well as the structure of historical wealth distribution evidence. After going through the country cases, we compile the series and study to what extent there are common patterns over time. Note that we focus on the twentieth and twenty-first centuries in the country-specific figures in order to ease inspection of the

[66] See Andreoni et al. (1998) and Slemrod and Yitzhaki (2002).

[67] Friedman et al. (2000) provide evidence supporting the idea that higher taxes also lead to better administration across a broad sample of countries as they find that higher taxes are associated with less unofficial activity.

[68] For 2001, the most recent year for which the IRS has final figures, the tax gap in the United States (i.e., the difference between taxes owed and taxes paid) was around 16%. Out of the 345 billion dollars that make up the tax gap, only about 4 billion were associated with estate and excise taxes.

trends in this era, whereas in the figures compiling several countries (Figures 7.19–7.21) we show the full set of observations stretching back to the nineteenth and eighteenth centuries.[69]

7.3.2.1 *Country-Specific Evidence*

7.3.2.1.1 Australia

A recent investigation of Australian wealth concentration since the beginning of the twentieth century is the one by Katic and Leigh (2013). The authors estimate top wealth shares using three different sources: estate tax returns, household surveys, and journalistic rich lists. The main emphasis is put on the first two, but the very recent trends can also be studied by putting the rich lists into context.

The earliest observation comes from a war wealth survey conducted in 1915 by the Commonwealth Bureau of Census and Statistics. From the 1950s up until the 1970s, tabulated estate tax returns were collected and adjusted by using inverse mortality multipliers adjusted for age, sex, and social status. From the 1980s onward, the authors again used wealth surveys, conducted by different entities, but complemented them by annual observations of wealth share of the superrich Australians published in the Australian magazine *Business Review Weekly*.

A common theme in all these sources is that they are not extensive in terms of coverage of wealth holders. With a few exceptions, only the very richest citizens are covered, and for this reason the only long-run time series coming out of the historical evidence are the wealth share of the top 1 and top 0.5 percentiles.

Figure 7.9 shows the trend in the Australian top wealth percentile share between 1915 and 2008. The share falls from almost 35% of total wealth during the First World War down to less than 15% in the early 1950s. Due to the lack of observations in between these dates, we cannot tell whether the fall came as a consequence of the immediate post-WWI turmoil, the crisis impact during the Great Depression of the 1930s, or the dramatic events during the Second World War and its aftermath. From the 1950s onward, the top percentile share has hovered at around a level of 10–15% of total wealth. Internationally, this is a very low wealth share, actually the lowest of all countries covered in this chapter. At this point, the reasons for the low Australian share have not been studied in detail.

7.3.2.1.2 Denmark

For Denmark, historical wealth concentration data exists from as early as 1789 and then more frequently during the twentieth century. The earliest observation comes from a comprehensive national wealth tax assessment in 1789, from which Soltow (1981a–c) collected a large individual sample of the gross wealth of households.[70] The next observation, however, comes over a century later at the time of the introduction of the modern

[69] This section is built partly on the cross-country analysis in Ohlsson et al. (2008).
[70] See Soltow (1981a–c, table 2).

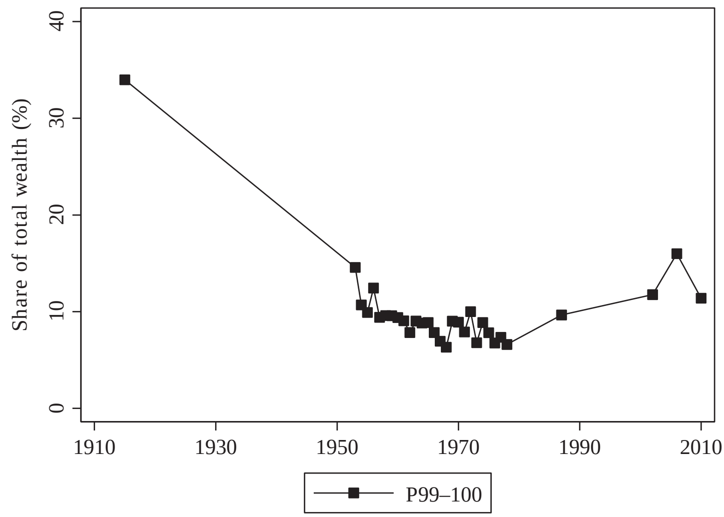


Figure 7.9 Wealth concentration in Australia, 1915–2010. *Source: See the Appendix for details about sources and data.*

wealth tax. For 1908–1925, Zeuthen (1928) lists tabulated wealth distributions (number of households and their wealth sums in different wealth size classes) for Danish households, adjusted to include those households with no taxable wealth. Similar tabulated wealth tax-based data are published in Bjerke (1956) for 1939, 1944, and 1949 and in various official statistical publications of Statistics Denmark for a few years thereafter until the wealth tax was abolished in 1997.[71]

Figure 7.10 shows the wealth shares of groups within the top decile between 1908 and 1996, while Figures 7.19 and 7.20 show the trends back to 1789. The lowest four percentiles (P90–95) exhibit a flat trend up to 1908 and thereafter double their share from 10% to 20% over the twentieth century. The next four percentiles (P95–99) lie constant between 25% and 30% of total wealth over the entire period, whereas the top percentile (P99–100) decreases significantly over the entire period, with particularly marked decreases after the two world wars. When looking at the very top of the distribution, the top 0.1 percentile (P99.9–100), there is no decrease at all up to 1915, but instead there is a dramatic drop by almost two-thirds of the wealth share between 1915 and 1925. Overall, the Danish wealth concentration decreased over the course of industrialization, and this continued throughout the twentieth century, although the development was not uniform at all times and across all groups.

[71] The estimates in 1995 and 1996 were constructed from evidence on only the tabulated number of wealth holders (families) and the total net wealth in the whole country. Supplementary Danish top wealth shares exist for the 1980s in Bentzen and Schmidt-Sørensen (1994), but unfortunately wealth size has been top-coded in their data, and the resulting estimates are not fully comparable with the other tax-based data.

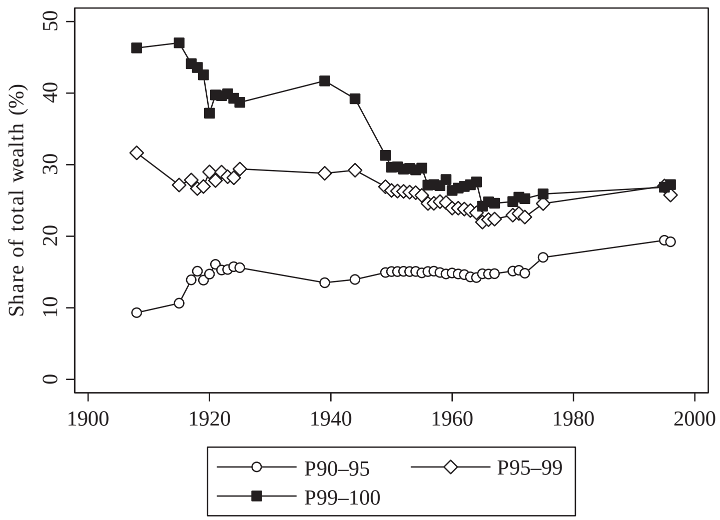


Figure 7.10 Wealth concentration in Denmark, 1908–1996. *Source: See the Appendix for details about sources and data.*

One way to understand the wealth compression of the Danish industrialization is to compare the identities of the Danish top wealth holders before and after the late nineteenth century. The dominant groups in the top of the wealth distribution in 1789 were owners of large agricultural estates. Soltow (1981a–c, p. 126) cited a historical source saying that "some 300 Danish landlords owned about 90 percent of the Danish soil." By contrast, in 1925, the group with the largest private fortunes was the brokers (*Veksellerere*) although landlords (*Godsejere*, *Proprietærer og Storforpagterere*) were still wealthy, both groups having more than 50 times larger average wealth than the country average.[72]

The drops in top wealth shares after the two world wars were partly associated with the sharply progressive wartime wealth taxes.[73] According to Bjerke (1956, p. 140), however, the fall after the Second World War was also largely due to new routines in the collection and valuation of wealth information of the tax authorities, which in particular made middle-class wealth more visible. Toward the end of the century, the wealth concentration continued declining up to the 1980s, largely due to increased share of the relatively equally distributed house-ownership in the total portfolio (Lavindkomstkommissionen, 1979, Chapter 5), but thereafter started to increase up to the mid-1990s.

7.3.2.1.3 Finland

Finland is another Nordic country for which wealth distribution data exist since the agrarian era and during most of the twentieth century. The country's industrialization

[72] The average net personal wealth in 1925 was Danish kronor (DKR) 6826 for all of Denmark, DKR 366,000 for brokers, and DKR 359,000 for large landlords (Zeuthen, 1928: 447).

[73] On the historical development of Danish wealth taxation, see Christensen (2003, pp. 8, 14).

came relatively late, in the interwar period, and even around the Second World War Finland was a predominantly agrarian economy focusing on forest industry and small-scale agriculture. Politically Finland was part of Sweden up until 1809, after which it came under Russian rule until 1917 when Finland ultimately gained independence (Eloranta et al., 2006).

Our estimates of the Finnish historical wealth distribution are essentially based on wealth tax statistics.[74] The earliest known observation of wealth distribution in Finland is 1800, coming from a wealth tax assessment levied in Sweden and Finland. Jutikkala (1953) and Soltow (1980) examined this assessment collecting a representative sample of the gross wealth of almost 2000 male household heads. The taxed households represented about one-third of the population, whereas the other two-thirds were exempt because they lacked a sufficient amount of personal taxable wealth. The next set of estimates comes from estate data in 1907–1909, 1914, and 1915 compiled and published by Statistics Finland.[75] We compute top wealth shares of the deceased but adjust these with respect to the likely difference between top wealth shares of the deceased and living populations using observed differentials in Sweden around the same time.[76] For the early twentieth century, we use Soltow's (1980) estimates from wealth tax assessments in 1922, 1926, and 1967. All these samples include adjustments for the share of the population without wealth on which no wealth tax was levied. Finally, we have wealth tax tabulations for the period 1987–2005 using net marketable wealth data retrieved directly from Statistics Finland.[77]

Figure 7.11 presents the evolution of wealth concentration in Finland from 1908 up to 2005, and Figures 7.19 and 7.20 show the trends back to 1800. The top decile held 46% of domestic net wealth in 1800, and its share peaked at 70% in 1909. Over the period the Finnish top percentile share exhibits an inverse-U shape, setting out at a relatively low share in 1800, which was doubled a century later in the years preceding the First World

[74] There are some previous studies, for example, Tuomala and Vilmunen (1988), who analyzed the distribution of wealth using tax data between 1968 and 1983, and Jäntti (2006) analyzing the developments during the latter 1980s and 1990s using mainly survey evidence. There is also a Finnish wealth survey run by Statistics Finland about twice per decade during the 1980s, but we do not use these data due to comparability issues with tax data as well as concerns about coverage in the wealth top. See Statistics Finland (2006, 2007). For example, the top decile wealth share in 1987 was as low as 35% in the survey, whereas it was 51% in the wealth tax statistics. The latter corresponds to a top percentile share of 16%, which is still in the lower region among Western countries.

[75] Statistics Finland (1911), table "Förmögenhetsförhållanden. A. Kvarlåtenskapsstatistik, 3. Arflåtna bon samt för dem uppburen stämpelskatt år 1909."

[76] In the Swedish investigation Finansdepartementet (1910), detailed calculations of wealth shares were made for both the deceased population (using estate shares) and the living population (using inverse mortality multiplier-adjusted evidence) for the years 1906–1908. It was found that the top wealth share of the deceased exceeded those for the living population by between 10% (for the top decile) and 40% (for the top 0.01 percentile).

[77] Data were submitted as a file by Statistics Finland showing taxable wealth by age and net wealth class in eight brackets from €5000 up to €500,000.

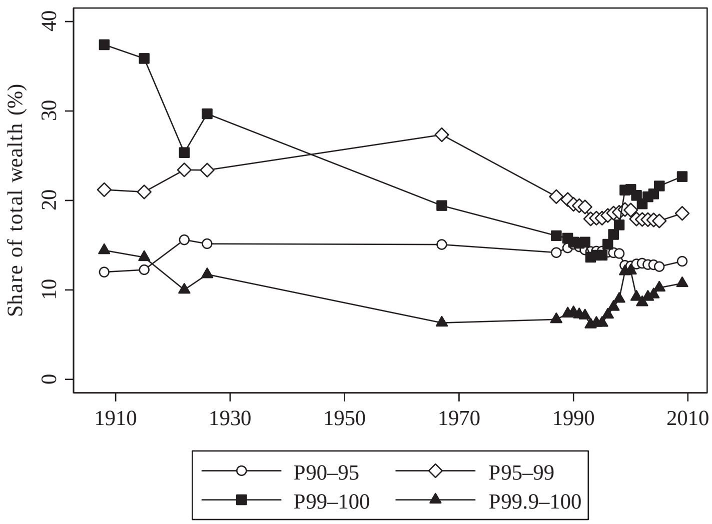


Figure 7.11 Wealth concentration in Finland, 1908–2009. *Source: See the Appendix for details about sources and data.*

War. The 1920s saw a strong reduction in the top percentile share, possibly due to the civil war taking place at this time. Later on during the twentieth century, the top percentile's share decreased further, reaching a global low in around 1990 when its share was less than 14% of total personal wealth. However, after this Finland experienced the IT-boom, led by the immense success of mobile phone producer Nokia, and the top percentile share increased swiftly during the 1990s and 2000s, reaching a level of 22% in 2005 (Eloranta et al., 2006). As for the rest of the top decile, the Finnish pattern is similar to that of most other countries studied here. The next four percentiles (P95–99) also experienced an inverse-U pattern, but peaked later, in the 1960s, after which its share started to decrease. The bottom half of the top decile hovered around 10–15% of total wealth.

Overall, the historical wealth concentration in Finland follows a pattern that looks very much like an inverse-U. The share of total wealth held by the rich (in the top percentile) increased during the nineteenth century and decreased during the twentieth century. The upper middle class (the rest of the top decile), however, did not change their relative position much during the two centuries covered. Also notable is the relatively low level of wealth concentration in Finland, especially in the year 1800 but also during the twentieth century.

7.3.2.1.4 France

The long-run evolution of French wealth inequality is particularly interesting to study given France's important role for Europe's economic and political development. Piketty et al. (2006), and later adjusted by Piketty (2014), presented new data on wealth concentration for Paris and France over almost 200 years from the Napoleonic era up to today. No previous study on any country has produced such long homogenous time series, offering complete coverage of wealth inequality over industrialization. The French wealth data comes from estate sizes collected in relation to an estate tax established in 1791 and maintained for more than two centuries. For every 10th year during 1807–1902, the authors manually collected all estate tax returns recorded in the city of Paris—Paris was chosen both for practical reasons but also because it hosted a disproportionally large share of the wealthy in France. Using summary statistics on the national level for the estate tax returns, the top Paris wealth shares were "extrapolated" to the national level. For the post-1902 period, tabulated estate size distributions published by French tax authorities were used.

Figure 7.12 shows the evolution of the wealth shares for some fractiles within the top wealth decile in France since 1900, while Figures 7.19 and 7.20 show the trends beginning in the early nineteenth century. The estimates are from the population of deceased, that is, directly from the estate tax returns, but comparisons with the equivalent wealth shares for the distribution of the living population (computed using estate multipliers) reveal practically identical trends and levels.[78] The figures show that wealth

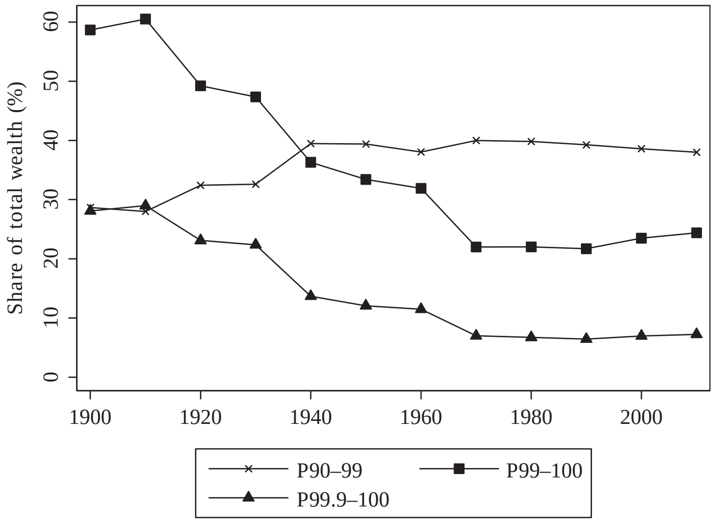


Figure 7.12 Wealth concentration in France, 1900–2010. *Source: See the Appendix for details about sources and data.*

[78] Using data in Piketty et al. (2004: tables A2 and A4) over top wealth shares for both the dead and living populations in Paris and France, it is evident that the trends in wealth shares over time is practically the same for all fractiles, and even the levels do not differ much, on average 0.4% for the top decile and 5.1% for the top percentile.

concentration increased significantly for the top 1 and 0.1 percentiles over the nineteenth century, first slowly up to the 1870s then more quickly until its peak at the eve of the First World War. By contrast, the two lower groups in the top decile are much less volatile during the period. The bottom half (P90–95) held about 9% of total wealth until the First World War when its share started to increase slowly until it had doubled by the 1980s. The next 4% (P95–99) stayed put on a level around 27% of total wealth throughout the period. These patterns suggest that the French industrialization, which took off around midcentury, greatly affected personal wealth. It did so already after a couple of decades, but only in the absolute top. This conclusion is further supported by two other observations. First, the composition of top wealth went from being dominated by real estate assets (mainly land and palaces) in the first half of the century to being dominated by financial assets (cash, stocks, and bonds), which were supposedly held by successful industrialists and their financiers. Second, over the same period the share of aristocrats among top wealth holders decreased from about 40% to about 10%.[79] From the First World War to the end of the Second World War, top wealth shares declined sharply, which according to Piketty (2003) is directly linked to the shocks to top capital holdings that inflation, bankruptcies, and destructions meant. The postwar era was quieter with regard to changes in the wealth concentration, although its decline continued most likely in relation to the increase of progressive taxation (Piketty et al., 2006).

7.3.2.1.5 The Netherlands

The Netherlands represents and interesting point of reference to the analysis of long-run trends in wealth inequality among Western economies. Although the Netherlands did not industrialize in the traditional sense until the middle of the nineteenth century, its economy was already developed due to its role in the expansion of global trade that started already in the sixteenth century. According to van Zanden (1998b), this may explain the apparent lack of increase in inequality following the Industrial Revolution. Although inequality grew during the preindustrial era due to high growth rates but stagnant real incomes, industrialization did not only boost fortunes of the wealthy but there was also an increased demand for all kinds of labor, skilled as well as unskilled.

The previous literature on historical wealth inequality in the Netherlands is relatively rich. Soltow (1998) and Vermaas et al. (1998) present a series of estimates of inherited wealth and housing inequality in different Dutch regions from the beginning of the nineteenth century. Unfortunately, defining a trend over the nineteenth century appears to be difficult. The only comparable information between 1808, 1880, and 1908 comes from inheritance tax records that cover inheritances to distant heirs, that is, not spouses and children. The data indicates a slight increase in inequality.[80]

[79] These facts are shown in Piketty et al. (2006: figures 4–6).
[80] See Vermaas et al. (1998, table 7.11, p. 167).

The most comprehensive longitudinal data are offered by the wealth tax statistics, which allow for an estimation of top wealth shares since 1894. The primary source of these observations is Wilterdink (1984), which presented a detailed account of the top vintile and groups within it for selected years between 1894 and 1974. The estimates stem from wealth tax records, showing the distribution among wealth tax units (mainly individuals), whereas the recent wealth survey data show the distribution among households. For the most recent years, Statistics Netherlands has compiled wealth-tax based distributions for the periods 1993–2000 and 2006–2011.[81]

Figure 7.13 shows the top wealth shares of the Netherlands from this year up to 2011. Wealth concentration was a high and stable around the turn of the century 1900. Thereafter the top percentile wealth share started decreasing. Both Wilterdink (1984) and van Zanden (1998a) highlighted the role of the geopolitical events, and these are clearly seen in the falls in top percentile shares during the two world wars and the depression of the 1930s. However, the researchers also emphasized the role of governmental redistribution, in particular the imposition of heavy wealth taxes after 1946 to finance the reconstruction after the war.

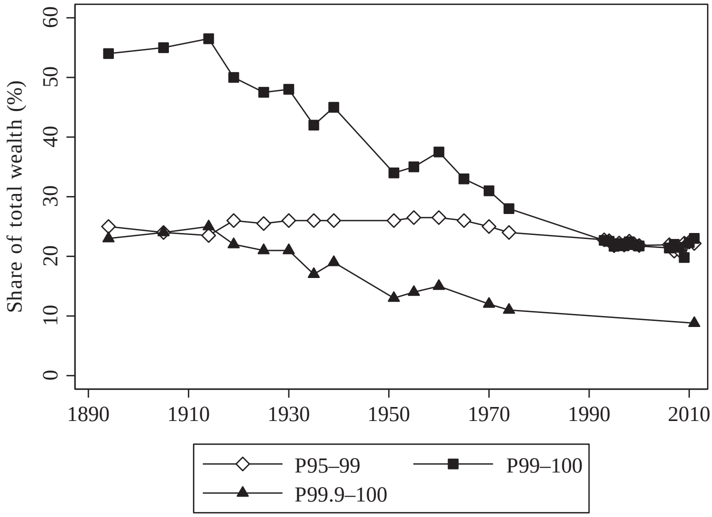


Figure 7.13 Wealth concentration in the Netherlands, 1894–2011. *Sources: See the Appendix for details about sources and data.*

[81] The data come from Statistics Netherlands and were kindly shared to us by Wiemer Salverda (see Salverda et al., 2013, pp. 47ff, and Statistics Netherlands, 2010, for further description of the data). A tax reform in 2001 introduced a slightly different mode of taxing wealth which reduces comparability of data. Furthermore, we only observe the top 5 percentiles in 1993–2003 (and assume their share of the top decile is the same during 2006–2011) and the top 0.1 percentile share in 2011.

7.3.2.1.6 Norway

Data on Norwegian wealth concentration come mostly from various kinds of wealth taxation. Overall, these data are perhaps the most uncertain presented in the entire chapter, and the estimates of top wealth shares presented in this chapter must therefore be interpreted cautiously. The first observation is from 1789, when the wealth tax assessment that also was launched in Denmark came into place (the two countries were in a political union at this time). As in Denmark, both real and financial assets were subject to taxation, including land, houses or farms, factories, livestock, mills, shop inventories, and financial instruments. Debts were not deducted, and hence the wealth concept is gross wealth.[82] Our second observation is from 1868, when the Norwegian government launched a national wealth tax assessment. Mohn (1873) presents totals for wealth and households and a tabulation of the wealth held by the top 0.27% (P99.73–100) of all households, including a detailed listing of the 15 overall largest fortunes.[83] For 1912, we use wealth tax returns from the taxation of 1913–1914 (exempting financial wealth) which are presented in tabulated form in Statistics Norway (1915b).[84] Similarly, for 1930 we use tabulated wealth distributions (number of wealth holders in wealth classes along with totals for wealth and tax units) presented in Statistics Norway (1934).

From 1948 onward, we use the tabulation of wealth holders and wealth sums in classes of net wealth published annually in the *Statistical Yearbook of Statistics Norway*. In the early 1980s the wealth statistics started being reporting for individual taxpayers instead of, as before, for households. To keep our series as consistent as possible, we attempted to convert the post-1982 observations from reflecting the individual distribution to reflect the household distribution using a listing of both types by Statistics Norway for the year of 1979.[85]

[82] We use Soltow's (1980) distributional estimates based on "males or families aged 26 and older," which is not identical to what is used for latter years and probably implies that the 1789 inequality should be adjusted upward to be more comparable.

[83] There is no information about whether it was the gross or net wealth that was taxed.

[84] We use tables of wealth holders in wealth classes in Statistics Norway (1915b: 20–21), corroborated by information about reference wealth and tax unit totals in Statistics Norway (1915a: 13f) and Kiær (1917: 22). The fact that financial assets were exempt in the Norwegian wealth taxation before 1922 is discussed in Statistics Norway (1934: 1).

[85] The *Statistical Yearbook of Norway* of 1981 tabulates the net wealth of both households (table 380: 316) and personal taxpayers (table 368: 306). In the latter case, however, we have no data on the sum of personal wealth of all wealth holders in each wealth class. We therefore insert the sums of wealth observed in household case into the individual case for the exact corresponding wealth classes. The comparison of wealth shares across these two distributions shows that the individual distribution produces shares that are 25%, 21%, 30%, 44%, and 60% higher than the household distribution for the top 10%, 5%, 1%, 0.1%, and 0.01% fractiles, respectively.

For the period since 1993, we use tabulated wealth distributions published on the Statistics Norway's Web site.[86] Somewhat ironically, the uncertainty about these data is perhaps largest because both asset coverage and valuations are highly problematic. For example, tax-assessed values of housing are heavily discounted and represent on average no more than a fifth of their true market value, with the discount being larger for more expensive dwellings (Epland and Kirkeberg, 2012). For this reason, household net tax-assessed wealth is negative for practically every Norwegian household. Furthermore, it is not obvious that the distributional trends in tax-assessed net assets are the same as those in market-valued assets if there are also trends in market-to-tax values of dwellings.

To shed some additional light on these matters, we refer to what we see as the most reliable estimate of the Norwegian net wealth distribution presented by Epland and Kirkeberg (2012). This investigation brings together a rich microdata material for 2009, carefully estimates market-valued assets and liabilities, and computes wealth inequality estimates. The study finds that the top wealth decile held about 53% and the top percentile about 21% of all net wealth (Epland and Kirkeberg, 2012, table 8). Interestingly, although the aforementioned tax-based tabulations of net wealth made no sense, the distribution of gross wealth seems less off the chart, producing for 2009 top shares of 54% for the top decile and 26% for the top percentile. For this reason, we use the time series pattern offered by the tabulated gross wealth of Statistics Norway and scale down the wealth shares to match the Epland–Kirkeberg reference level of 2009.

Altogether, the Norwegian long-run wealth concentration estimates are thus highly problematic in several respects. Looking at the overall trend in wealth concentration, however, it appears to be relatively robust to variations in some of our assumptions, and it does not deviate much from the long-run inequality trends observed in other countries.

Figure 7.14 presents the trends in Norwegian wealth concentration between 1912 and 2002, while the trends back to 1789 are shown in Figures 7.19 and 7.20. The top wealth decile is broken up into the bottom 5% (P90–95) of wealth holders, the next 4% (P95–99), the top percentile, as well as the top 0.1 percentile. Norway's top wealth holders experienced quite different trends in their relative positions over the period. As for the bottom 5% of the top decile, its share decreases between 1789 and 1912 and then jumps up sharply between 1912 and 1930 to land on a fairly stable (though slowly declining) level thereafter. The wealth share of the next 4% exhibits an inverse-U shaped pattern, increasing sometime in the nineteenth century (we do not know exactly when due

[86] See www.ssb.no/statistikkbanken (2013-10-28). For the period 1993–1999, see table "Tabell: 08575: Fordeling av skattepliktig brutto- og nettoformue for busette personar 17 år og eldre, etter talet på personar og gjennomsnitt i kroner (avslutta serie)." For the period 2000–2011, see the table "Tabell: 08532: Fordeling av skattepliktig brutto- og nettoformue for bosatte personer 17 år og eldre, etter antall personer og gjennomsnitt."

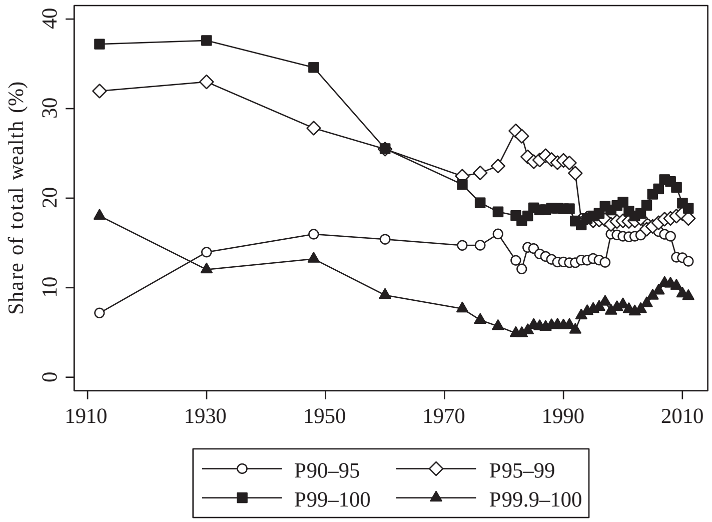


Figure 7.14 Wealth concentration in Norway, 1912–2011. *Sources: See the Appendix for details about sources and data.*

to a lack of data), peaking in 1930 and then declining almost monotonically over the rest of the twentieth century. Finally, the share of the top wealth percentile decreases significantly between 1789 and 1868, both years predating Norway's industrialization period. The share then goes up to slightly in 1912 only to start decreasing again. The most dramatic falls occur in the postwar period, with the top percentile dropping from 34.6% to 17.5% during 1948–1983 and the top 0.1 percentile going from 13.2% to 5.7% over the same period. In the 1990s, there is a rapid recovery, which may be related to the oil fortunes being built up in recent times and to the rise in world stock markets prices that also produced a rise in the top income shares over this period (Aaberge and Atkinson, 2010). The sizeable increase between 1997 and 1998 can also be explained by a change in the Norwegian tax laws specifying an increase in the assessed values of corporate stock on personal tax returns.[87]

Despite the somewhat disparate trends among Norway's top wealth holders and underlying problems with the Norwegian wealth tax-based data series, the evidence presented in Figures 7.14, 7.19, and 7.20 nonetheless corresponds relatively well with what one could expect given the economic and political history of Norway over this period. The Norwegian economy was badly hit by the economic crisis after the Napoleonic wars, and there was a shift in the political power from the great landlords and landed

[87] The tax-assessed values of stocks were raised in 1998 for stocks listed at the Oslo Stock Exchange from 75% to 100% of the market value and for nonlisted stocks from 30% to 65% of a stipulated market value.

nobility to a class of civil servants.[88] When merchant shipping expanded in the world after 1850, Norwegian shipowners and manufacturers experienced a tremendous economic boost. When looking at the average wealth of various occupations in 1868 listed in Mohn (1873: 24), the four richest groups were manufacturers (having 160 times the country average household wealth), merchants (124 times), shipowners (96 times), and civil servants (87 times). Half a century later, in 1930, a similar comparison between the wealth of top occupations groups and the country average was made (Statistics Norway, 1934, p. 6), and only shipowners had kept the distance to the rest of the population (having 119 times the country average wealth), whereas merchants (22 times) and manufacturers (19 times) had lost wealth relative to the average.

7.3.2.1.7 Sweden

In a recent study, Roine and Waldenström (2009) compiled available evidence of historical wealth distribution data for Sweden to construct a homogenous series of top wealth shares from the time of the industrial takeoff in the late nineteenth century up to the early 2000s.[89] The primary basis for these series was wealth tax statistics published in various sources, including censuses and special public investigations by tax authorities or the Ministry of Finance. The concept of wealth in these sources is typically net wealth in tax-assessed values. However, these data were complemented by estate tax material originally presented by Ohlsson et al. (2008) for a few points in time: 1873–1877, 1906–1908, 1954/55, 1967, and 2002–2003. A striking resemblance between wealth tax and estate tax data emerges regarding the patterns over the twentieth century. In addition to these sources, there is also an early observation of Swedish gross wealth inequality in 1800 using evidence collected by Soltow (1985) from a national tax assessment.[90] This observation comes from a wealth census that was carried out in 1800 and describes the gross wealth distribution for the population of males aged 20 and older.[91]

Figure 7.15 shows the evolution of top wealth shares since 1908 while Figures 7.19 and 7.20 depict the trends over the past two centuries. Looking first at the pattern over the nineteenth century, our observations indicate a relatively stable wealth distribution, which by today's standards was very unequal. As there are no observations between 1800 and 1873 (or actually 1908), there is little that can be said about the nineteenth-century

[88] Historical account taken from the section on Norway's history during "The Napoleonic Wars and the 19th Century" in *Encyclopædia Britannica Online*.

[89] Prior to the study by Roine and Waldenström (2009) and series presented in Ohlsson et al. (2008), the long-run evolution of Swedish wealth inequality was also studied by Spånt (1979, 1982).

[90] The observant reader notes that Finland experienced a similar tax in the same year. These taxes were part of the same assessments because Finland still was part of Sweden during this period. Our analysis, however, is confined to Sweden's current borders.

[91] About one-third of the males were wealth holders, and we adjust for the remaining two-thirds when computing the inequality estimates (using data in Soltow, 1985, table 5, p. 18).

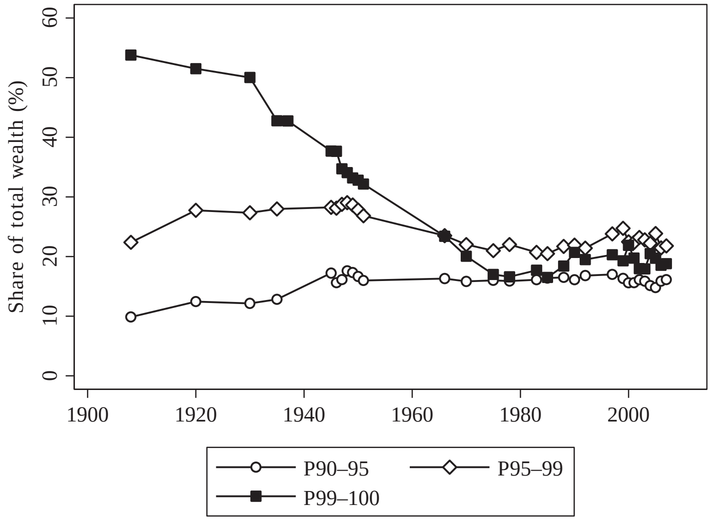


Figure 7.15 Wealth concentration in Sweden, 1908–2007. *Sources: See the Appendix for details about sources and data.*

development. However, Soltow (1989a) made attempt to do so using public reports about the amount of citizens in four specified social classes ("destitute," "poor," "moderately rich," and "rich") between 1805 and 1855 and some other sources of property distribution. His main conclusion regarding the wealth inequality trend is that overall inequality seems to have decreased over this period and all the way up to the twentieth century.[92] Soltow admits, however, that his calculations do not exclude the possibility that the top 1 or 2 wealth percentiles may have actually increased their share of total private wealth.

Over the twentieth century the picture is much clearer. We are able to use multiple sources that overlap in time, and even though there is still uncertainty about the levels over time, the trends seem relatively certain. The long-run trend in wealth concentration in Sweden over the twentieth century is that the top decile has seen its wealth share drop substantially, from around 90% in the early decades of the century, to around 53% around 1980, and then recovering slightly to a level around 60% in recent years. In the bottom half of the Swedish wealth distribution there is a considerable share of households holding negative net wealth, a fact that appears to be partly due to widespread state loans for college studies but partly also because several important assets, for example, condominiums and private and public pension savings, are not fully covered in the official wealth statistics.

[92] See Soltow (1989b, tables 1 and 2, pp. 49–53).

Looking just at this general trend is, however, incomplete if one is to really comprehend the evolution of wealth concentration. Decomposing the top decile as shown in Figure 7.15, we see that the majority of the top decile actually experiences substantial gains in wealth shares over the first half of the century. The overall drop in the top decile share is explained by such dramatic decreases in the top percentile share that this outweighs the increase for the lower groups in the top decile. In the period 1950–1980 the entire top 5 percentile experiences declines in wealth shares, but the decrease is larger for the top percentile, and after 1980 the trend is again the same for both groups but now the gains in wealth shares are somewhat larger for the top percentile.

How can we account for these developments? Focusing first on the decreases in the very top of the distribution over the first half of the century we note that most of the decrease takes place between 1930 and 1950, with the sharpest falls in the early 1930s—a time of financial turbulence and in particular the collapse of the Kreuger company empire—and just after the Second World War.[93] The period after 1945 was a time when many of the reforms discussed in the 1930s, but put on hold by the war, were expected to happen, and politically the Communist Party gained ground forcing the Social Democratic Party to move to the left.[94] In particular, the progressive taxes that had been pushed up during the war remained high and also affected wealth holdings as Sweden had a joint income and wealth tax until 1948.

The main reason for the decreasing share in the very top is, however, likely to be the increasing share for the lower part of the top decile, and this, in turn, is likely to be increased wealth accumulation among relatively well-paid individuals. After 1945 the trend of increased accumulation of wealth continues down the distribution. Over the next 30 years the most important change is the increased share of owner-occupied housing in total wealth, which increases from being 17% of all wealth to 45% in 1975 and remains around that in 1997 when adding owner-occupied apartments and houses and vacations homes (Roine and Waldenström, 2009). Even if this type of wealth was far from evenly accumulated across the distribution, it accrued to relatively large groups in the distribution causing wealth concentration to keep falling. Today about half of all households in Sweden own their homes.[95] Over the past decades fluctuations in wealth shares depend largely on movements in real estate prices and share prices. Increases in the

[93] Although Sweden was not as affected by the Great Depression as many other countries, the so-called Kreuger Crash in 1932, the bankruptcy of Ivar Kruger's industrial empire, led to major loses of wealth in Sweden.

[94] See, for example, Steinmo (1993).

[95] A specific feature of the Swedish wealth distribution in recent decades is the large share of negative net wealth holders, almost a third of the adult population. There are several factors explaining this characteristic, including widespread state loans for higher education and an underreporting of important assets such as condominiums and private and public insurance savings (see further, e.g., Cowell, 2013; Jansson and Johansson, 2000).

former have a tendency to push up the share of the upper half of the distribution at the expense of the very top causing inequality to go down, whereas increases in share prices make the very top share larger due to share ownership still being very concentrated causing inequality to increase.[96] In the year 1997 the top percentile in the wealth distribution owns 62% of all privately held shares, and the top 5% holds 90%.[97]

7.3.2.1.8 Switzerland

Data on the Swiss wealth concentration are based on wealth tax returns compiled by tax authorities for disparate years between 1913 and 1997 and analyzed by Dell et al. (2007). The Swiss wealth tax was levied on a highly irregular basis, and the authors have therefore spliced several different point estimates from local as well as federal estimates to get a roughly continuous series for the whole country.

Figure 7.16 depicts top wealth shares within the Swiss top wealth decile over the twentieth century. In stark contrast to the other countries surveyed in this study, wealth concentration in Switzerland appears to have been basically constant throughout the period. The wealth shares at the top of the distribution have decreased, but the movements are small compared to all other countries studied.[98] This does not only refer to the

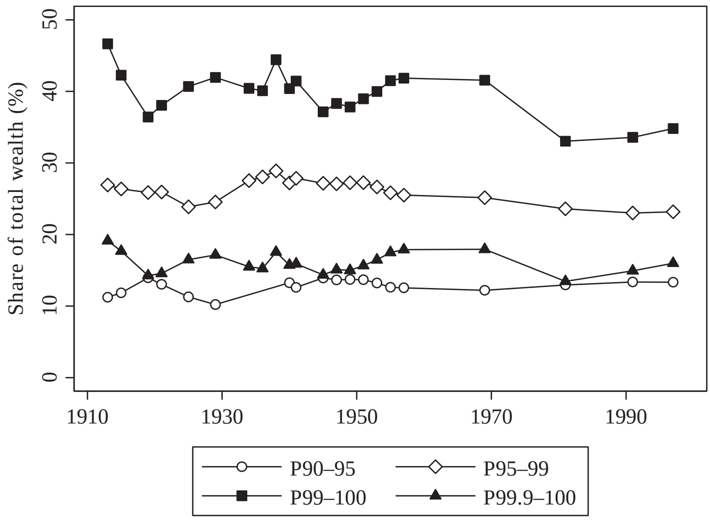


Figure 7.16 Wealth concentration in Switzerland, 1913–1996. *Sources: See the Appendix for details about sources and data.*

96 In their study of the role of capital gains Roine and Waldenström (2012) discussed some aggregate asset developments in Sweden since the 1980s. This indicates that the surge in financial asset values has been much larger than overall real estate values.

97 Jansson and Johansson (2000, pp. 38–40).

98 A simple trend regression yields small but significant negative coefficients.

top decile vis-à-vis the rest of the population, but perhaps most strikingly also to the concentration of wealth within the top. The highest percentile and the top 0.1 percentile have not gained or lost considerably compared the bottom 9% of the top decile, except for some short-run fluctuations.

Accounting for this long-term stability of Swiss wealth inequality is not easy. One possibility is the country's relatively low level of wealth taxation, which suggests a low rate of redistribution and small effects on the incentives to accumulate new wealth. The twentieth-century experience with high taxes on wealth and inheritance appears to have contributed to the low top income and wealth shares in a number of countries, as we discuss elsewhere in this chapter. However, the fact that Switzerland stayed out of both world wars cannot alone account for the stable wealth distribution; Sweden also escaped both world wars does not share the Swiss pattern. In any case, the Swiss top wealth share series seriously questions the hypothesis that significant economic development always lead to a lower level of wealth inequality over time for reasons of either redistribution or simply relatively quicker accumulation of household wealth among the middle class.

7.3.2.1.9 United Kingdom

There are a number of estimates of the wealth concentration in the United Kingdom dating back to the country's industrialization in the middle of the eighteenth century. Prior to 1900, data on wealth distribution are less homogenous and emanate from scattered samples of probate records and occasional tax assessments (see Lindert, 1986, 2000; Soltow, 1981a–c). It was not until the Inland Revenue Statistics started publishing compilations of estate tax returns after the First World War that the series are fully reliable (see Atkinson and Harrison, 1978; Atkinson et al., 1989).[99] Still there are some notable breaks in the series. For example, the geographical unit of analysis changes over time, with pre-Second World War numbers almost always being England and Wales, whereas the post-war ones reflect all of the United Kingdom. Data in Atkinson et al. (1989, table 1) show, however, that the differences between these entities are fairly small. More important, the tax authority changed some of its methods to compute top wealth shares leading to large breaks in the time series around the Second World War, in 1960, and around 1980. Among the important changes were lowered age cutoffs, different treatment of life insurance policies and valuation of consumer durables, and also more careful collection routines of the tax authorities.[100]

[99] The Inland Revenue actually started publishing estate tax data in 1896, but it was not divided by age, which precludes estimation of the wealth distribution using the mortality multiplier method. Note also that the mortality multipliers used by the Inland Revenue were based on age only up until 1923 when they were based on both age and sex.

[100] See further Atkinson and Harrison (1978, Chapter 6) and Atkinson et al. (1989) for an extensive discussion of these breaks.

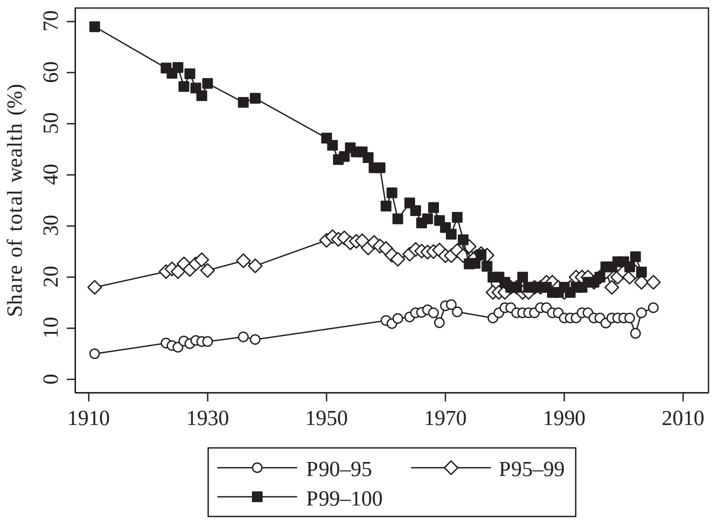


Figure 7.17 Wealth concentration in the United Kingdom, 1911–2005. *Notes and sources: England and Wales up to 1960, Great Britain thereafter. See the Appendix for details about sources and data.*

When England industrialized in the second half of the eighteenth century, the buildup of personal wealth also changed. Looking at the overall wealth concentration since 1911 in Figure 7.17 and Figures 7.19 and 7.20 for the period back to 1740 it is evident that there is great heterogeneity within the top 5 percentiles of the distribution.[101] Apparently, wealth concentration at the very top increased, while, by contrast, the wealth share of the next four percentiles saw its wealth share decline during the same period. Using supplementary evidence on personal wealth, Lindert (1986, 2000) shows that wealth gaps were indeed increasing in the absolute top during the nineteenth century, with large landlords and merchants on the winning side. At the same time, Lindert points out that the middle-class (in this case those between the 60th and 95th wealth percentiles) were also building up a stock of personal wealth, and this is probably what is causing the drop in the share of the next 4% in Figure 7.17.

After the First World War, the pattern was the reversed. While the top percentile wealth share dropped dramatically from almost 70% of total wealth in 1913 to less than 20% in 1980, the share of the next four percentiles remained stable and even gained relative to the rest of the population. Atkinson et al. (1989a,b) argued that this development was driven by several factors, but that the evolution of share prices, the ratio of consumer durables, and owner-occupied housing (so-called popular wealth) to the value of other wealth were the most important ones. According to the most recent statistics from the Inland Revenue, the top percentile's share increased by about one-third between 1990 and 2003, but this increase has not yet been explained by researchers.

[101] The reader should keep in mind that this figure, and several others in this study, contains spliced series coming from different sources, which naturally may impede the degree of homogeneity over time.

Possibly, it reflects the surge in share prices following the financial market deregulation of the 1980s as the financial wealth are most concentrated to the absolute top of the wealth distribution.[102]

7.3.2.1.10 United States

The historical development of wealth concentration in the United States has been extensively studied by economists and historians, and estimates are available back to the time of the American Revolution. In this study, we combine different pieces of evidence to create a long and relatively homogenous series of wealth inequality. As acknowledged by previous scholars, there are several problems concerning consistency over time, which has spurred some controversy over both definitions of data and conclusions drawn. For these reasons, we compare some of the complementary series using different sources and wealth definitions to get an idea of how large these problems may be.

Our focus is the evolution of U.S. top wealth shares from colonial times to the present day. The main series refer to the distribution of net wealth among households, and for these we show wealth shares of fractiles within the entire top decile. Still the figure also presents the top percentile shares in the adult distribution for which there are rich, annual data available over especially the twentieth century. The top wealth shares for the household distribution prior to 1900 are few but important as they determine our notion of the link between industrialization and inequality in the United States. There has been some disagreement over the pre-1900 inequality trends, with some scholars arguing that preindustrial U.S. inequality was high and that inequality was basically stable during the nineteenth century (e.g., Soltow, 1971, 1989), whereas others have argued that U.S. wealth inequality increased markedly between the Revolution and the latter half of the nineteenth century (e.g., Lindert, 2000; Williamson and Lindert, 1980a,b). In this chapter, we use the observations reported by Lindert (2000). These are essentially the estimates from the seminal contributions of Alice Hanson Jones (see, e.g., Jones, 1970, 1972, 1980), which included adjustments to add unfree men and women to the reference total population.

The available evidence for the twentieth century is more unified, with long-run series being based on a combination of estate tax returns and survey data (see, e.g., Lampman, 1962; Smith, 1984; Wolff and Marley, 1989). We use the compilation of those sources by Wolff (1996) for the period up to 1958, and for the period thereafter we use the survey data from the SCF and its forerunners presented by Kennickell (2009, 2011).[103] For the

[102] This is a stylized fact that is true for many developed countries (see, e.g., the overview of "stylized facts" in Davies and Shorrocks, 2000).

[103] Wolff (2012) also used SCF data to compute a series of U.S. wealth concentration since 1962, but his series deviates from those of Kennickell (2009, 2011). Wolff explained this with his exclusion of consumer durables from the wealth concept, motivated by the fact that these are neither easily marketable nor included in the national accounts-based definition of household wealth. In this chapter, we use Kennickell's series as they match the earlier evidence from U.S. surveys and estates, which consistently included consumer durables among household assets.

adult population, our preferred estimate for 1774 is from Lindert (2000).[104] For the nineteenth century, there are unfortunately only gross wealth estimates for the adult population (see Lindert, 2000), and therefore the next evidence is for the years 1916–2000 provided by Kopczuk and Saez (2004) using mortality multiplier-adjusted federal estate tax returns.

Figure 7.18 shows the results for the period since 1916 and Figures 7.19 and 7.20 for the period back to 1774. Beginning with the two top percentile series, they appear to be inversely U-shaped over the period, with wealth shares increasing slowly between the late eighteenth and the mid-nineteenth centuries but then much faster between 1860 and 1929, when they more than doubled. The long-run pattern of the lower 9% of the top wealth decile, however, exhibits stable or even decreasing shares of total wealth (although based on rather few observations). This inequality increase in the absolute top coincides with the industrialization era in the United States around the mid-nineteenth century. Although the few pre-First World War estimates are uncertain, their basic message is supported by researchers using other sources. For example, Rosenbloom and Stutes (2008) also found in their cross-sectional individual analysis of the 1870 census that regions with a relatively high share of its workforce in manufacturing had relatively more unequal wealth distributions (see also Moehling and Steckel, 2001). Another anecdotal piece of evidence in support for a linkage between industrialization and increased

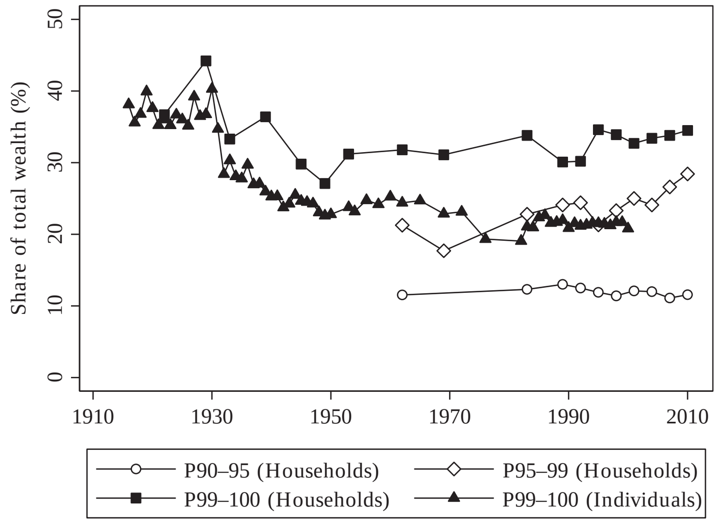


Figure 7.18 Wealth concentration in the United States, 1916–2010. *Sources: "Households" and "Individuals" refer to different wealth holder populations. See the Appendix for details about sources and data.*

[104] As Lindert (2000, footnote to table 3) noted, this estimate deviates slightly from that of Shammas (1993) because the latter also includes the wealth of British residents living in the U.S. colonies.

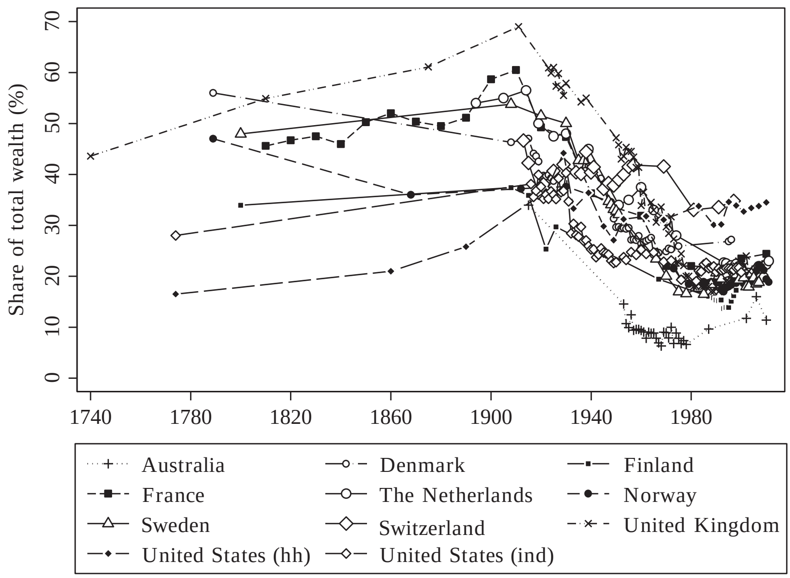


Figure 7.19 Wealth concentration in 10 countries, 1740–2011. *Sources: Graph shows top percentile (P99–100) wealth shares for all countries. See the Appendix for details about sources and data.*

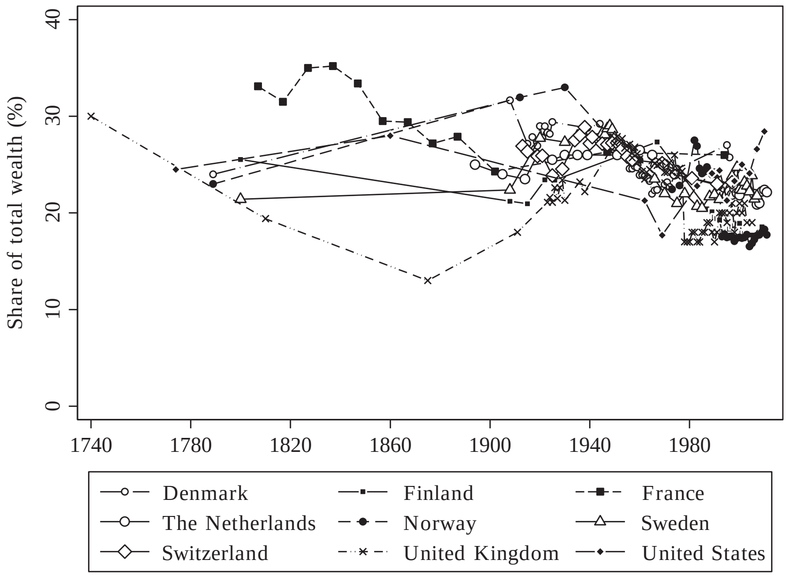


Figure 7.20 Wealth share of the "next four percentiles" (P95–99) in nine countries. *Sources: See the Appendix for details about sources and data.*

inequality is that the 15 richest Americans in 1915 were industrialists from the oil, steel, and railroad industries and their financiers from the financial sector.[105]

The twentieth-century development in Figure 7.18 suggests that wealth concentration peaked just before the Great Depression when the financial holdings of the rich were highly valued on the markets. In the depression years, however, top wealth shares plummeted as stocks lost almost two-thirds of their real values. Kopczuk and Saez (2004) showed that corporate equity represented more than half of the net wealth of the top 0.1 percentile wealth holders in 1929. Another contributing factor to wealth compression was surely the redistributive policies in the New Deal. After the Second World War, the top percentile wealth shares remained low until the 1980s, when the top household percentile's share increased significantly, peaking around mid–late 1990s and then to decline somewhat in 2001. By contrast, the top adult percentile wealth share from the estate series in Kopczuk and Saez (2004) exhibits no such increase, which is surprising given that this period also saw a well-documented surge in U.S. top incomes (Piketty and Saez, 2003). Whether the difference in trends between the household and adult distributions reflects inconsistencies in the data or some deeper dissimilarity in the relation between income and wealth accumulation remains to be examined by future research.

7.3.2.2 Cross-Country Trends in Long-Run Wealth Concentration

Earlier we presented a compilation of recent as well as some new evidence on the long-run evolution of wealth inequality in 10 Western countries: Australia, Denmark, Finland, France, the Netherlands, Norway, Sweden, Switzerland, the United Kingdom, and the United States. As we have already pointed out, the quality of these data differs substantially across countries and in some cases even within single countries over time. Like many previous researchers, we have attempted to adjust the series to make them consistent and comparable over time, but naturally some problems remain. Still, we have tried to classify the series (countries) into different quality levels to run the analysis on more homogenous subsets, and those exercises do not produce any notably different conclusions with respect to the long-run trends in wealth concentration.[106]

Figure 7.19 shows the top wealth percentile in each of these countries for various periods during 1740–2011. Furthermore, Figure 7.20 contrasts the trends in the top

[105] See the listing of the top 20 fortunes in 1915 by De Long (1996).

[106] For example, there are some countries for which the data are quite composite in terms of data sources, coverage of assets, and so on (e.g., Australia, the Netherlands, and Norway), but removing them from the analysis does not change the overall pictures reported. It is more difficult to adjust further for qualitative breaks over time within countries, and both the Norwegian and the U.K. series during the twentieth century contain some important break points in data definitions and quality. However, even if we had acted differently when connecting the segments separated by break points for these countries, we would still have observed a marked long-run trend toward wealth compression in both these countries.

percentile against those in the next four percentiles (P95–99). Even though great caution should be taken when comparing these series, we still believe that some conclusions can be drawn about the developments of wealth inequality in these countries over the past 200 years.

Two broad conclusions can be drawn from the series as summarized in Table 7.5. First, the evidence does not unambiguously support the idea that wealth inequality increases in the early stages of industrialization. Looking at the development of the wealth share of the top percentile among the countries analyzed here, the Nordic observations indicate fairly stable inequality levels over the initial stages of industrialization (i.e., in the late nineteenth century). The U.K. series (England and Wales) exhibits clearly increasing wealth shares for the top percentile in the period of the two industrial revolutions (1740–1911), as do the U.S. and French series over the nineteenth century. For the Netherlands, the evidence is less certain, indicating either a flat or a slightly increasing nineteenth-century trend (van Zanden, 1998b; Vermaas et al., 1998). Overall this suggests that going from a rural to an industrial society, with entirely new stocks and types of wealth being created, may, but does not necessarily, give rise to a large increase in wealth concentration. It also suggests that—just as in the case with income inequality series—carefully studying smaller fractiles of the distribution is necessary to get a more complete picture of the development.

Second, although the series do not suggest a clear common pattern over the nineteenth century when industrialization took place (first in the United Kingdom, later in the United States, France, and the Netherlands and toward the end of the century in the Nordic countries), the development over the twentieth century seems more uniform. Top wealth shares have decreased sharply in just about all countries studied in this

Table 7.5 Wealth inequality trends across eras, 10 Western countries

	From industrial takeoff to the First World War		From the First World War to 2000	
	P99–100 (Top 1%)	**P95–99**	**P99–100 (Top 1%)**	**P95–99**
Australia	–	–	Decrease	–
Denmark	Decrease	Flat	Decrease	Flat
Finland	Flat	Flat	Decrease	Flat
France	Increase	Flat	Decrease	Flat
Netherlands	Flat?	Flat?	Decrease	Flat
Norway	Flat	Increase	Decrease	Decrease
Sweden	Flat	Flat	Decrease	Flat
Switzerland	–	–	Flat	Flat
United Kingdom	Increase	Decrease	Decrease	Flat
United States	Increase	Flat?	Flat/Decrease	Flat?

Notes: The nineteenth century inequality trends for the Netherlands are not observed directly, but various sources indicate that there was little increase in inequality during the Dutch industrialization since the middle of the century (see Section 7.3.2.1.5 on Netherlands).

chapter with the exception of Switzerland and possibly also the United States, where the fall has been small, but where the level also was not as high historically as in most European countries. The magnitude of the decrease seems to be that the top percentile lost its share of total wealth by about a factor of 2 on average (from around 40–50% in the beginning of the century to around 20–25% today). It also seems that the lowest point in most countries was around 1980 and that the top percentile wealth share has increased in most countries after that. Interestingly, the wealth share of the next 4 percentiles (P95–99) does not display any strong indications of a decreasing trend. Indeed, there are periods of notable equalization also affecting this wealth fractile, but over the course of the entire century Table 7.5 clearly highlights that this moderately rich group sustained its share of total wealth. This said, there were likely replacements between economic groups and types of actors over time (as also suggested by the country case studies earlier), indicating that the cross-sectional evidence also needs to be complemented by evidence about mobility within the distribution.[107]

Similar to the analysis of long-run top income shares, we can make a closer examination of the evolution of wealth concentration expressed in terms of wealth shares of the very top groups within the larger top group. This approach results in a slightly different measure of inequality as it looks at the inequality *within the top* of the wealth distribution and not overall inequality. As some theories are especially concerned with widening gap among the rich, investigating inequality among the wealthy can make sense.[108] Furthermore, estimating the reference total wealth held by the full population is associated with potential error. Applying the shares-within-shares measure by dividing the top wealth percentile by the top wealth decile, P99–100/P90–100, we land at a ratio that effectively eliminates the reference total.[109]

Figure 7.21 depicts the evolution of wealth concentration using the shares-within-shares estimate. Two countries drop of out the picture (Australia and the Netherlands) due to a lack of long-run data on the top wealth decile, and there are also fewer observations for the countries still in the comparison. Still the patterns confirm our previous findings. The equalization of the twentieth century is clearly observed except for in the Swiss (and possibly the United States) cases. As for the nineteenth-century development, the picture gets a bit blurry, largely due to a lack of data. The Nordic countries exhibit similar inequality trends as given earlier: rising in Finland and Sweden but falling in Denmark and

[107] See, for example, the study by Edlund and Kopczuk (2009), which found that the share of women in the U.S. wealth top fluctuated and that this indicates changes in the relative importance of dynastic versus entrepreneurial wealth.

[108] There are several theories that in various forms imply an advantage for the very top of the distribution, for example, the superstar model of Rosen (1981). See Section 7.4 in this chapter for more.

[109] Similar to the result found for top incomes, for top wealth percentile P99–100 $= W_{\mathrm{Top1}}/W_{\mathrm{All}}$ (with $W=$ Wealth) and top wealth decile P90–100 $= W_{\mathrm{Top10}}/W_{\mathrm{All}}$, we get P99–100/P90–100 $= (W_{\mathrm{Top1}}/W_{\mathrm{All}})/(W_{\mathrm{Top10}}/W_{\mathrm{All}}) = W_{\mathrm{Top1}}/W_{\mathrm{Top10}}$.

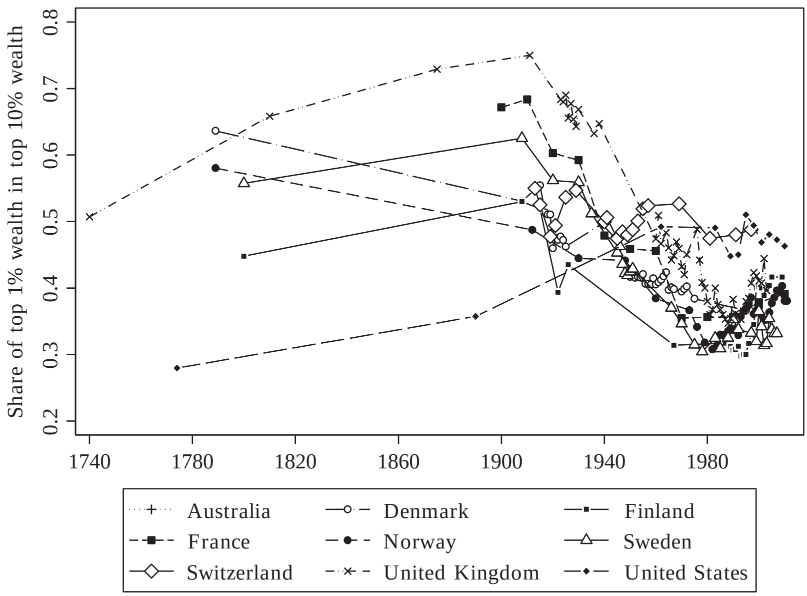


Figure 7.21 Shares-within-shares (P99–100/P90–100), nine countries, 1740–2011. *Notes and sources: The shares-within-shares measure is computed by dividing the top wealth percentile (P99–100) by the top wealth decile (P90–100). The resulting measure eliminates the (often separately constructed) reference wealth total from the equation and thereby offers a robustness check of the overall trends. However, the measure also provides a metric of the wealth inequality within the top. See the Appendix for details about sources and data.*

Norway. France also looks quite similar as when the actual top shares are examined. Overall this implies that, notwithstanding the variations, most of the long-run wealth inequality trends are driven not by the changes of the very top in relation to those just below, but by the change of the entire wealth top in relation to the rest of the population.

7.3.3 The Composition of Wealth

Up until this point, the analysis has dealt primarily with the distribution of total net wealth. However, the composition of wealth across asset types (and debts) also matters to wealth inequality trends just as the composition of labor and capital incomes was shown in the previous sections to matter for the trends in income inequality. Unfortunately, when it comes to the historical evidence about wealth composition across the wealth distribution, we know almost nothing. As for the aggregate composition of private wealth, we know more thanks to both old and more recent evidence.[110] These data show that agricultural assets practically vanished over the course of the two past centuries. Private housing, by contrast, increased its share of total national wealth from one-fifth in

[110] See Chapter 15 by Thomas Piketty and Gabriel Zucman in this Handbook, and the references therein.

the nineteenth and early twentieth centuries to three-fifths today, largely corroborating the previously documented postwar rise in "popular wealth," dwellings, and consumer goods, among the broad layers of the population.[111]

Among the few studies that contain evidence on the wealth composition across different groups of wealth holders, less than a handful offer some kind of historical evidence. One stylized fact that seems to hold regardless of time period, however, is that financial assets in general, and corporate securities in particular, are consistently more important in the portfolios of the rich than of the rest of the population. For example, Kennickell (2009) and Cowell (2013) showed that the share of basically financial assets in the top wealth decile was higher than for the population as a whole (except for savings, which are more important for middle-class households between the median and the 90th wealth percentile). Kopczuk and Saez (2004) showed that the share of corporate stock in the portfolios of top 0.5 percentile U.S. wealth holders (using estate tax data) was between 40% and 60% during 1916–2000 and that this was strictly higher than for the whole population (using national wealth estimates).[112]

In their study of trends in French wealth concentration, Piketty et al. (2006) documented similar patterns for France over the nineteenth century. Specifically, they looked at the share of personal estate, which includes all nonreal assets, in total assets and found that its share was higher among the richest in the top 0.1 percentile than among the intermediately rich in the rest of the top wealth decile. It was, however, also very high among the broad layers population (the bottom nine wealth deciles). The authors explained this U-shaped pattern by the fact "that real estate is a middle class asset: the poor are too poor to own land or buildings; what little they have is in furniture, cash, or other moveables. In contrast, the rich hold most of their wealth in stocks and bonds" (p. 244).

Altogether, the historical evidence on the composition of wealth across the distribution suggests that housing wealth is more important in the portfolios of the broader population, whereas financial assets dominate the portfolios of the rich. Furthermore, the new long-run evidence on aggregate wealth of the private sector shows that housing wealth became more important in total national wealth after the Second World War, and this fact probably explains a large part of the documented wealth compression witnessed in many Western economies during this period.

7.3.4 Concluding Discussion: What Do the Long-Run Wealth Inequality Trends Tell Us?

What then can we say about the relationship between wealth concentration and economic development based on the data reported in this chapter? Can one talk about

[111] See Piketty and Zucman (2013, appendix table A18) and, for an early observation of the postwar rise in popular wealth, Atkinson and Harrison (1978).

[112] The relative difference has varied notably from at least double as high (from the 1930s to the 1980s) to less than 10% (in year 2000).

common patterns across countries over the development path or are there mainly a set of disparate country-specific histories? Have initial wealth inequalities been amplified or reduced? Taking stock of the series shown here suggests that industrialization was not unambiguously accompanied by increasing wealth inequality. Although inequality did indeed increase in the United Kingdom, the United States, and France, it probably did not change much in the Netherlands, Finland, Norway, and Sweden and even decreased a little in Denmark. Noting that the countries in the first group all were large, central economies that were early to industrialize, whereas the Netherlands and the Nordic countries were smaller economies that industrialized later, may hold clues to the different experiences but it does not change the fact that industrialization did not increase wealth concentration everywhere.

The experience over the twentieth century appears to be much more homogenous. As the countries continued to develop, top wealth concentration also dropped substantially. Looking at the details of the pattern by which different fractiles gain wealth shares indicates that this drop was due to a gradual process of wealth spreading in the population—confirming the role of increasing "popular wealth" identified in, for example, Atkinson and Harrison (1978). In a sense this pattern is consistent with a Kuznets-type process in which inequality eventually decreases as the whole economy becomes developed. However, this development was probably not driven by the kind of process suggested by Kuznets, but mainly by other factors such as political interventions and exogenous shocks. Piketty et al. (2006) argued that it primarily was adverse shocks to top wealth during the period 1914–1945 related with wartime shocks that decreased French wealth inequality and that the subsequent introduction of redistributive policies that prevented them from recovering. Piketty (2011) and Piketty and Zucman (2014) emphasized that the wartime shocks to capital were only to a limited extent the consequence of outright destructions of factories, constructions, or infrastructure, instead pointing at the importance of capital taxation and regulation. A similar explanation is given by Kopczuk and Saez (2004) for the United States.[113] This reasoning has been supported by the fact that Switzerland, which did not take part in either of the wars, exhibits rather stable top wealth shares. Our data on Sweden, which also did not participate in any of the world wars, shows an example of equalization taking place without decreases in top wealth shares driven by exogenous shocks. Even though events such as the Kreuger Crash in 1932 hit top wealth holders in Sweden as well, this does not explain the entire drop. Policy may, at least in Sweden, have played a more active role in equalizing wealth than merely holding back the creation of new fortunes after the Second World War. Suggesting that rising taxation and increased redistribution have been important for the decline of wealth inequality is also consistent with the largest drops taking place in the Scandinavian countries as well as with the smaller decline in Switzerland, with its smaller government.

[113] Scheve and Stasavage (2010, 2012) showed the increased taxation of capital and high incomes was indeed political developments associated with wartime events.

Altogether the data presented here suggest that (a) there was a mixed impact of industrialization and (b) in later stages, after countries became industrial, significant wealth holding spread to wider groups, bringing wealth inequality down. In terms of the often discussed inverse U-shape over the path of development, the first upward part does not seem to be present everywhere, whereas the later stage decrease in inequality does fit all countries we have studied. An important addition to this characterization is that this analogy misses an important point that is present in the series. Whereas the inverse U-shape suggests that the distribution of wealth starts at some level in a nonindustrialized society, then rises, and later returns to the same level of inequality, all our series indicate that development has unambiguously lowered wealth concentration. The proper characterization of wealth inequality over the path of development hence seems to be that, so far, it follows an inverse J-shape with wealth being more equally distributed today than before industrialization started. The direction of future inequality remains to be seen.

7.4. DETERMINANTS OF LONG-RUN TRENDS IN INEQUALITY

How can we understand the trends in the distribution of income and wealth outlined in the previous sections? Do the series systematically relate to other developments in society that have been suggested to influence inequality, and if so, in what ways? How can we connect the observed long-run trends to existing theories about inequality? These are questions that we address in this section.

A number of facts that are likely to be important have already been noted in the previous sections along with the characterization of the trends. A first point is that an understanding of the development involves both wage and capital income, and thereby the dynamics are at least in part jointly determined by the distributions of income and wealth. For example, the drop in top shares over the first half of the twentieth century was largely a result of decreased capital incomes in the top, which in turn was largely driven by decreasing wealth shares in this group. High marginal tax rates in the decades after World War II made recovery difficult and caused top shares to decrease even further. We will explicitly look at these explanations in Section 7.4.2.

When it comes to the increase in top income shares since around 1980 this seems to be primarily related to increasing top wages, especially in the United States, but increasing capital incomes in the top also play a role in many countries (such as Sweden), especially after around 1990.[114] The increased earnings dispersion is often attributed to aspects of globalization and technological change. Many have pointed to technological change being skill-biased, usually equated with an increasing education premium, as a possible reason for increasing wage differences. But skill-biased technological change does not

[114] See Chapter 8 for a detailed view of inequality developments since 1970, where the increasing role of capital income in the top is also noted.

automatically lead to increasing wages of the "skilled." The impact on wage dispersion depends on several things, such as the structure of the production function and the change in the supply of educated workers.[115] Unless the dynamics in "the race" between technology and education are made explicit, skill-biased technological change can be consistent with any number of "education premium" profiles over time.[116] Furthermore, even if one were to focus on a version of the model where the increased demand for skills actually lead to increasing wage dispersion, it is difficult to see how this would explain that so much of the increase is concentrated in a relatively small top group. To account for such increases within the top it seems necessary to find something that distinguishes a small fraction of the "skilled" from others who are equally educated (at least in terms of observables). Examples of such explanations include a number of so-called super-star theories, where technology and globalization disproportionately have benefitted those who—for various reasons—are most in demand in their field. Others have emphasized the possible role of changing norms. Some of the theories that have been put forth to understand the rise in top earnings over the past decades will be the subject of Section 7.4.3.

Finally, in Section 7.4.4 we present an overview of some recent econometric evidence on correlations over the long run. These regressions do not constitute tests of any particular theory but nevertheless give some insights as to what relationships seem to be present in the data.

We will begin the current section, however, with offering a broad overview of major events and societal trends that have been suggested to influence the distribution of income and how these correspond to our long-run pattern of top income shares. We will also discuss what the new series imply for our understanding of the Kuznets curve. Our conclusion is that, even if some broad trends are consistent with proposed broad explanations, we cannot distinguish between alternatives just based on looking at how inequality has developed. Instead we need to look more carefully at developments in different parts of the distribution, at the source of income, and in particular on how income and wealth relate to each other and also relate all these aspects to predictions from theory.

[115] A technology that makes skilled workers more productive *decreases* the wage per unit of skill but *increases* relative demand of skilled workers. This alone can drive the wage of skilled workers both up and down. In addition the response (and the speed of it) in the supply of skills will determine the movement in relative wages of the skilled and unskilled.

[116] See Atkinson (2008a) for an explanation of the textbook model and a thorough discussion of other aspects that need to be considered. On the importance of the dynamics, he noted, "Surprisingly, the dynamics of wage differentials seem to have been little discussed in the literature of recent years. Yet, there is good historical precedent. In 1959, Arrow and Capron published a paper on dynamic shortage and price rises, with an application to the then shortage of engineers and scientists—an application that seems of contemporary relevance" (Atkinson, 2008a, p. 10).

7.4.1 A First Look at Inequality Trends, Structural Changes, and Shocks

What is the relationship between top income shares and the broad societal changes that have been hypothesized as affecting distributional outcomes? How well do the basic patterns match? Next we will discuss inequality developments in relation to trends in globalization, technological breakthroughs that have altered production in society (often referred to as general purpose technologies), inequality in relation to wars and shocks to the economy, and finally, inequality in relation to economic growth.

Globalization has been suggested to affect inequality in a number of ways. Classical trade theory in the spirit of Eli F. Heckscher and Bertil Ohlin has a clear prediction for inequality: In countries relatively abundant in skilled labor and capital (developed countries), inequality increases, whereas the reverse is true in (low-skill) labor abundant developing countries, where instead inequality goes down.[117] Modern trade theory is less clear-cut. Although some effects, like the gains to the largest most productive firms (in models like Melitz, 2003; Melitz and Ottaviano, 2008) seem to suggest increasing returns in the top, others have pointed to globalization being most beneficial for the top *and* the bottom, while hurting individuals in the middle of the distribution (e.g., Leamer, 2007; Venables, 2008). Yet others have pointed to the possibilities of efficiency gains from globalization being so large that these effects can compensate losses from, for instance, offshoring (Grossman and Rossi-Hansberg, 2008).

Looking at the inequality developments over what has been labeled as different waves of globalization, the first wave (1870–1914) coincides with flat or increasing inequality, followed by decreasing inequality in the antiglobalization period (1914–1950).[118] As most countries for which we have data belong to the relatively skill and capital abundant, this could be seen as in line with theoretical predictions.[119] The second wave of globalization is harder to reconcile. In 1950–1980 measures of globalization (trade flows/GDP, foreign capital as share of GDP) clearly increase while inequality clearly decreases. There are some obvious counterarguments to this. First, one may argue that the level of globalization was not yet sufficiently high for the predicted effects to show, but second,

[117] These effects remain across the many versions of Heckscher–Ohlin type models. See, for example, Wood (1994) for a summary of the basic arguments.

[118] Clearly the definition of what constitutes a period of globalization is somewhat arbitrary. Most authors seem to agree that there was a globalization period before 1914, though there is disagreement on when it started. It is also commonly accepted that the period between 1914 and 1945 was an era of increased protectionism characterized by drastically smaller economic flows between countries. This was gradually reversed after 1945. To emphasize the difference between the intensity in globalization some refer to the period between 1945 and 1980 as a second wave (with gradually increasing globalization) as different from the period after 1980, when globalization really took off. See Lindert and Williamson (2003) and World Bank (2002) for details on different views on periods of globalization.

[119] Note, however, the important point made by Williamson (2006) that the effects depend on the *relative* abundance more than on a country being rich or poor. Inequality developed in opposite directions in the periphery depending on the labor abundance/scarcity.

one could also point out that this was a period when most of capital flows and trade was between developed countries.[120] If one places the start of the recent era of increased globalization around 1980 instead, the pattern is more promising as the period thereafter is characterized by increasing inequality. A problem is, of course, that during this period inequality has been increasing in developing countries too, counter to the basic Heckscher–Ohlin model.[121]

What about innovations leading to skill-biased technological change? Such shifts play a major role in the large literature trying to explain recent changes in the earnings distribution. Models building on Tinbergen's (1974, 1975) seminal work suggest that the returns to skills are determined by a race between education (creating a supply of skilled workers) and technology (implicitly technology that complements skills). Technological change pushes in the direction of increased wage differences between skilled and unskilled, unless education keeps up and creates an increased supply of skilled workers that keeps down the wage differences. Goldin and Katz (2008) bring much of this work together in a unified framework. Acemoglu and Autor (2012, 2013) give overviews of much of this literature and also claim that these models have been empirically successful in accounting for recent wage dispersion mainly based on U.S. data (e.g., Autor et al., 2006; Katz and Autor, 1999; Katz and Murphy, 1992).

But, as already pointed out, skill-biased technological change does not automatically result in increased wage differences (and even less automatically in increased inequality). Even in the simplest model the outcome depends on the speed of the supply response, and depending on the relative shifts in demand and supply of skills, the resulting wage differential between the groups can look different. In particular, this means that even if countries are affected by the same technological change, the impact on the wage distribution may look very different depending on how responsive countries are in terms of improving the skills in the population. See Atkinson (2008a,b) for more details and additional caveats to the simple model.[122]

Another historical aspect of technological change, noted, for example, by Caselli (1999), is that it has not always been skill biased. Indeed, some of the technological advances in the late eighteenth and early nineteenth centuries replaced, rather than complemented, skilled artisans and increased the productivity of low-skilled workers (Mokyr, 1990). Later advances, such as the electrification of industry in the late nineteenth century, seem to have been more skill biased. Firms using more electricity paid workers higher wages, workers were more educated, and these firms had higher capital ratios (Goldin and Katz, 1998). But soon thereafter the introduction of the assembly line at

[120] This observation is indeed the basis for much of the developments in trade theory since the late 1970s.
[121] See Freeman (2011) for more on the relationship between globalization and inequality.
[122] Also see Atkinson (1999) for an early critique of overly simplified versions of skill-biased explanations.

Ford's Highland Park facility in 1913 seems to be another technology shift that increased the relative productivity of unskilled workers.

If (and this is a potentially big "if") one accepts that technology shifts that are skill biased always lead to increasing inequality, and vice versa for deskilling technological change, then the basic historical pattern looks promising. The skill-biased electrification coincides with increasing or at least unchanged inequality, the introduction of the assembly line coincides with the start of the long decline in inequality, and the recent ICT revolution starting in the 1970s and 1980s also happens at the time when inequality turns up again. But obviously this does not mean that we can conclude anything about the relationships. In addition to the many assumptions needed, there are some other factors that are potentially problematic for a simplistic story of technological change driving common patterns of inequality. One is that technological changes do not take place everywhere at the same time. Comin and Mestieri (2013) give an overview of technology adoption lags and show that these can be very long. Second, given what we know about the role of capital in explaining the declining inequality in the first half of the twentieth century this seems separate from an explanation emphasizing returns to skills and an increasing earnings dispersion. Third, and perhaps most important, an explanation that focuses on the returns to higher education surely includes everyone in at least the top decile group. As such it cannot explain the large changes *within* the top and the fact that much of the recent increase has been limited to the income growth in the top percentile rather than a broader top group.

Shocks in the form of wars and major financial crises constitute yet another broad category of explanations. As already noted in previous sections, these events certainly seem to have had an impact on top shares, especially in some countries, and in particular on capital incomes. The exact degree to which the equalizations following after the wars were due to outright destruction of capital owned by the wealthy or whether taxes and regulations redistributed wealth and increased overall socioeconomic mobility seem to have varied across countries. We discuss this issue further later.

Another broad topic concerns the relationship between inequality and economic growth. The crudest possible illustration of this could be done by dividing history since 1870 into four broad periods based on the overall inequality trends and calculating the average yearly growth rate over these. Starting in 1870 the average growth rate until today is 1.82% for the countries in the sample. Dividing this period into four subperiods—1870–1914, characterized by increasing (or unchanged) inequality; 1914–1950 characterized by rapidly decreasing inequality; 1950–1980, when inequality continued to decline but at a slower rate; and finally the period 1980–today, when inequality has been increasing—we can examine the average growth rate in each of these periods. It turns out that only one of these subperiods has an average growth rate higher than the long-run average 1.82%, namely 1950–1980, when average growth was 3.18%. This period is characterized by falling top income shares. The lowest growth rates are in the late 1800s and early 1900s when inequality was relatively flat (or rising), and growth rates in-between can be found both in the past 30 years 1980–2010 when inequality has increased, and in the period 1920–1950, when top shares declined. Based on this, it is certainly hard to see any clear secular (bivariate) relationship between inequality and growth.

7.4.1.1 What About the Kuznets Curve?

Despite Lindert's (2000, p. 173) urge to the profession to "move onto explorations that proceed directly to the task of explaining any episodic movement, without bothering to relate it to the Kuznets curve," we find it difficult to avoid discussing the Kuznets curve in this chapter. In the end we will, however, perhaps even more clearly thanks to the new evidence we have, come to the same conclusion.[123]

In its crudest interpretation, equating the Kuznets curve with the question, "Is it true that inequality first increases and then decreases as a country develops?" the answer must clearly be "No." The fact that the broad pattern of decreasing inequality up until around 1980 has been followed by a sharp increase in some countries (but not all) clearly shows a pattern that is not consistent with inequality following an inverse U-shape, nor is it consistent with changes in inequality being the same across countries at similar levels of development. When testing the hypothesis on broad cross-country samples and in particular on developing countries, the evidence is mixed and inconclusive (Kanbur, 2000). With a broader interpretation it could be argued that increasing inequality in recent decades is, in fact, the start of a new Kuznets curve. The technological development starting in the 1970s constitutes the start of a shift, not from agriculture to industry as in Kuznets' original story, but from traditional industry to an ICT-intensive sector that initially rewards a small part of the population, but eventually will spread, bringing inequality down. This idea would, under a number of assumptions, fit the general pattern better.

But even in its broader interpretation, a number of aspects do not fit the Kuznets curve hypothesis. First, when it comes to the first half of the twentieth century, a main finding of the recent top income literature is that most of the decline is a capital income phenomenon. Even if there was a continuous decline in the share of workers in agriculture and a large rural–urban migration, their impact on wage inequality was small; low-wage rural workers were mostly replaced by low-wage urban workers (see the discussion in Section 7.2.3.3 and also Piketty, 2006, 2007). The inequality decline was, as evidenced by the timing of the fall, the source of income, and the concentration of the fall to the top percent group, due to shocks to wealth holders from the wars, the depression, and anticapital policies.[124] Second, the recent increase since around 1980 has the problem that it

[123] Also see Piketty (2006, 2007) for an account of how the recent top income literature impacts the view of the Kuznets curve.

[124] As Piketty (2011, p. 10) put it: "In effect, the 1914–1945 political and military shocks generated an unprecedented wave of anti-capital policies, which had a much larger impact on private wealth than the wars themselves."

does not fit the predicted earnings dynamics within the distribution. As an increasing number become skilled, the difference within the top should decrease, not increase as seems to be the case.[125]

Taken together the preceding suggests that there is no mechanical relationship between inequality and industrialization or technological change. It is no more unavoidable that inequality increases at early stages of introducing new technology, than it is automatic that inequality eventually goes down. The Kuznets curve conjecture has indeed played an enormous role in shaping the research on long-run changes of inequality, but the recent research has made it even clearer that it is time to follow Lindert's (2000) suggestion to look at long-run changes "without bothering to relate it to the Kuznets curve." In a way, part of the evidence suggests that other aspects also pointed to by Kuznets (1955) deserve more attention. After all, he formulated the famous Kuznets curve as a suggestion of how to explain what he saw as a puzzle of decreasing inequality. It was a puzzle because what he saw as the more obvious forces at play suggested that inequality be *increasing* in the countries he looked at: "There are at least two groups of forces in the long-term operation of developed countries that make for increasing inequality in the distribution of income..." (Kuznets, 1955, p. 7). The first of these forces had to do with the interplay between the concentration of savings, the impact this would have over time on capital incomes, and income inequality. In forces that could counter such a mechanical increase of concentration he pointed to political decisions and taxation. These are all features that play a major role in potential explanations that we will look at in the next section.

7.4.2 Combining Wage Earnings and Wealth

The relationship between savings, income, and wealth discussed by Kuznets (1955) pointed to the need for a theory where individuals both work and receive income from capital, in different proportions. Such a theory was developed by Meade (1964). In his framework individual wealth holdings grow along with savings, s, and returns to capital, r, but diminishes across generations as the wealth is divided among a growing population that is $1+n$ times larger in every period. If $sr \geq n$ wealth grows without limit but if $sr < n$ then the division of wealth exceeds the growth of wealth, and wealth holdings converge to being a multiple of earnings.

Stiglitz (1969) embedded Meade's framework in a general equilibrium model. Assuming that individual output is $f(k)$, with k being capital per worker, a competitive rate of return, r, being the same for everyone and equal to $f'(k)$, and population grows at rate n, aggregate capital converges to a steady-state level where $sf(k)/k = n$. This in turn implies that $sr < n$ so that in equilibrium division dominates growth of capital, and eventually the only thing determining wealth inequality is differences in earned income. This result, however, hinges on estates being divided equally. If one instead assumes that

[125] See Atkinson (2008a,b, p. 13) for more on this point.

wealth is inherited by one child (as with primogeniture), so that wealth is not divided, long-run wealth inequality is compatible with the $sr < n$.

Furthermore, the resulting distribution will have a Pareto upper tail with Pareto coefficient $\alpha = \frac{\ln(1+n)}{\ln(1+sr(1+t))}$, where $sr(1-t)$ is rate of accumulation out of wealth net of taxes, t (see Atkinson and Harrison, 1978, Chapter 8). This also suggests an empirical specification where we regress $1/\alpha$ on $sr(1-t)/n$.[126] However, even if primogeniture has been applied in the past (and still exists) the assumption of inheritance not being divided is implausible. In fact, today it is not even legal in most European countries. But there are several other assumptions that can be changed with the result that wealth inequality remains in equilibrium even if earnings are the same. Bourguignon (1981) shows that with a convex savings function there can be multiple locally stable equilibria, and with imperfect capital markets individuals with initially low wealth can be stuck in a "poverty trap." Introducing stochastic elements allows for the possibility of escaping such a trap, but also introduces a new source of wealth inequality. Benhabib and Bisin (2007) showed how introducing an idiosyncratic rate of return results in a Pareto distribution for wealth that depends on capital income as well as inheritance taxes.[127]

Besides providing the first model to treat individual incomes as jointly determined by income and wealth, Meade (1964) also provided a basis for studying the joint impact of changes in wealth concentration and changing factor shares on the income distribution. To illustrate using the top percentile group, their share of total income can be broken down into one part based on earnings and one part originating from wealth holdings in the following manner:

Share of top percentile = (proportion of earned income) × (share of top wage earnings percentile) × (alignment coefficient for earnings) + (proportion of capital income) × (share of top capital income percentile) × (alignment coefficient for capital income).

The alignment coefficient for earnings is the share in earnings of top percentile of income recipients divided by share of top percentile of wage earners and defined correspondingly for capital income. This captures the extent to which top wage earners and capital income recipients are also in the top of the total income distribution. In a class model where workers and capitalists are totally separate groups, there is zero alignment; workers have only earnings and capitalists only capital income. In a life cycle savings model with no inheritance, on the other hand, the same individuals inhabit the top of both earnings and capital income distributions, and the alignment is unity.

[126] Approximating $\ln(1+n)$ by n, and $\ln(1+sr(1-t))$ by $sr(1-t)$; see Atkinson et al. (2011), p. 58. See also Atkinson (2007).

[127] Piketty (2000) provided an overview of models of persistent wealth inequality. Recently, much important work has also been done on optimal taxation in models where income stems not only from individual actions over a lifetime but also from bequest flows from previous generations. In general, this changes many of the standard results in important ways. See Piketty and Saez (2013a,b). See also Chapter 15 by Thomas Piketty and Gabriel Zucman in this Handbook, and the references therein.

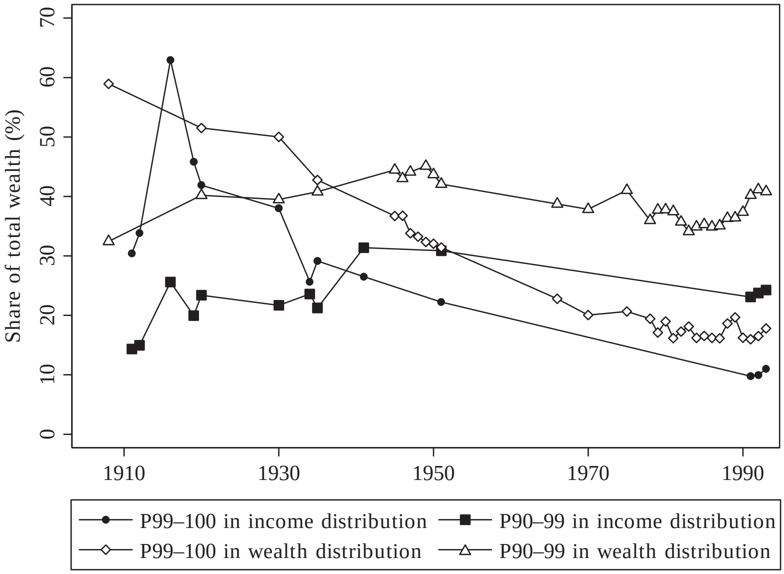


Figure 7.22 Wealth in top income and wealth fractiles in Sweden, 1908–2004. *Notes and sources: See main text and Roine and Waldenström (2008) for further details.*

Using this labor–capital decomposition it is, in principle, possible to attribute shifts in top income shares to shifts in top earnings shares, top capital income shares, and factor shares. A practical empirical problem, however, is that in most cases we lack data on the cross distributions over long periods of time. Roine and Waldenström (2008), studying Sweden, is an exception. Thanks to a particular form of combined income and wealth tax it is possible to calculate the distribution of wealth ranked both by wealth and total income.[128] Figure 7.22 shows that the share of total wealth when ranked by total income is somewhat lower than when ranked by wealth, but the two series are highly correlated, suggesting that there is significant overlap between the two distributions.

7.4.2.1 Explaining the Drop over the First Half of the Twentieth Century: Wealth Shocks and the Cumulative Effects of Taxes

Even if it is in most cases not possible to get a complete picture of the alignment of the distribution of earnings, capital income, and total income, a key feature of the top income data is the possibility to decompose income according to source. And, as already discussed in Section 7.2, it is clear that the drop in inequality in the first half of the twentieth century is mainly a capital income phenomenon. Combining what we know about the

[128] Between 1910 and 1948 Sweden had a form of wealth tax according to which a share of individual wealth holdings (initially 1/60, later 1/100) was added onto other incomes. The tabulations of incomes therefore also contain wealth amounts by income groups. In addition, for a few years wealth and income tax data can be matched on an individual level.

composition of the drop according to income source (almost entirely capital income driven), the timing (in most countries concentrated to wartime and the Great Depression periods), and the development of wealth concentration (large decreases in wealth concentration), declining capital incomes among top earners constitute the main explanation for declining top shares.

It is interesting to note that this development came about even in counties not immediately exposed to all of the great shocks of the twentieth century. Sweden is a case in point. The world wars did affect the Swedish economy, but the country never participated directly in either of them, and looking at details in and around these periods it is clear that they did not constitute immediate shocks to Swedish wealth holders. If single events are to be pointed out, the economic crises in the early 1920s, the indirect effects of the Great Depression, which hit Sweden in 1931, and, in particular, the dramatic collapse of the industrial empire controlled by the Swedish industrialist Ivar Kreuger (the "Kreuger Crash") in 1932, stand out as being most important. These are, however, not sufficient to explain the drop in top shares in Sweden. Instead, a trend of decreasing share of capital in value added corresponds well to the declining top income shares. Policy, especially sharp increases in top tax rates, also stand out as important for explaining especially the drop just after the Second World War.

The general picture thus seems to be that macroshocks explain most of the drop, but there is also a role for a shift in policy and probably also in an economy-wide shift in the balance between returns to capital and labor.[129]

Assuming that we are satisfied with the explanation of why top shares dropped, we then face the challenge to explain why they did not recover in the decades after the Second World War, but rather continued to decline. Here a key factor seems to be the high rates of marginal taxation facing the top. The long-run evolution of statutory top marginal taxes is shown in Figure 7.23. As a broad generalization, top rates started to increase rapidly in the 1930s and reached high levels in many countries during and just after the Second World War.[130] As shown in Piketty (2001a,b, 2003), the combined effect of

[129] The fact that macroshocks and financial crises led to decreased top shares in this period is not the same as saying that this is the expected outcome of financial crises in general. When wealth concentration is high, a sharp decline in its value translates to decreasing incomes from it. But if top incomes are primarily based on earnings (as in many countries in recent decades) the effect need not be large. Furthermore, it is also possible that in developed financial markets rich wealth holders can protect themselves against shocks using various instruments, or even by altering the rules of the game in their favor (as argued by, e.g., Reich, 2010). See Jenkins et al. (2013) for an overview of the recent financial crises pointing to varying effects on inequality across countries.

[130] Note, however, that statutory taxes and actual taxes paid by top income earners are not the same. In particular, during the 1950s and 1960s available evidence suggests that top rates were only paid well above the P99 level (see Roine et al., 2009 and references therein). Scheve and Stasavage (2010, 2012) argued that the increased wartime taxation of the rich can be related to a political process of equal sacrifices, where the wars forced the masses to put up with their poverty and in return the rich were forced to put up with their wealth.

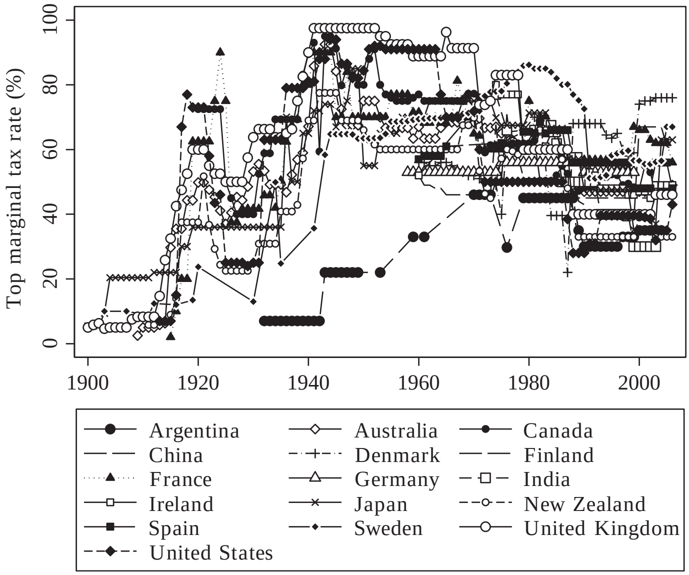


Figure 7.23 Trends in top marginal tax rates, 1900–2006. Sources*: See main text.*

shocks to capital holdings and high marginal tax rates is that recovery takes long. Unless adjustments to consumption are not made, current consumption levels can be sustained for some time by running down wealth holdings even further, but this decreases future income from wealth even more. An important point to note is that in these processes the short-run effect from taxes looks small. It is the cumulative effect over time that is important.[131]

How much do taxes impede capital accumulation and the recovery of top income shares? Just to illustrate the order of magnitude, assume a simple case with two groups of income earners; a top group that derives half their income from capital (the rate of return is assumed to be 5%) and the other half from wages, whereas the rest only have a wage income. Initially the income share of the top group is 15% of all income, and their consumption is such that their capital stock remains unchanged. These assumptions are, of course, not calibrated to fit any particular economy, but the numbers fit an approximate representation of the relationship between the top percentile and the rest of the population, both in terms of the importance of capital (with a broad interpretation) and the income share around World War II.

The combined effect of a tax increases from 30% to 60% and a shock that leaves the top group with 0.7 times initial wealth causes a gradual decline of both the capital income share (from 50% to 37% in 5 years and to 30% in 10 years) and total income share for the

[131] Roine et al. (2009) also showed that the cumulative effects of the relatively small short-run impact of taxes found in their econometric analysis are consistent with much larger effects over time.

top group (from 15 to 12.3 in 5 years, down to 11.1 in 10 years) with wages and consumption being unchanged. Despite the stylized nature of the setup, these magnitudes are reasonable when looking at data from the 1930s and following the Second World War. In the scenario with no changes to consumption wealth is eventually used up, and capital income goes to zero. Altering consumption too little makes the process longer but in the end the result is the same, whereas a sufficient adjustment allows for accumulation over time.[132]

7.4.3 Explaining Increasing Top Wages: Executive Compensation, Superstar Effects, and the Possibility of Changing Norms

Although shocks to capital income combined with the cumulative impact of high marginal taxes are important in explaining the development between the First World War and around 1980, something else is needed to account for the increasing inequality since. This is especially true for understanding changes in top earnings, most visible in the United States, but also clear in many other countries (see Atkinson, 2008a, Chapter 4).[133] As discussed in Section 7.4.1, it has been argued that increasing wage dispersion can be explained by theories of skill-biased technological change (Acemoglu and Autor, 2012, 2013) but also that these theories have some problems. In particular, it is hard to see how an increased advantage of the skilled, typically equated with the well-educated, squares with increased inequality not just in general but also increased inequality *within* the top group.

There are several strands of literature that give insights into why the top of the earnings distribution may behave differently from the rest and what factors may govern compensation of top performers. This includes theories of determination of earnings in hierarchical organizations, tournament theory, and superstar effects. Research based on these ideas, and sometimes combinations of them, has tried to account for the sharp increase in top wages in recent decades.

In models first developed by Simon (1957) and Lydall (1968) pay is related to the number of individuals supervised and to a (constant) pay increase at each step in the hierarchy. They assume that, first, at every level of the organization individuals supervise a constant number of people at the level below and, second, that the salary of these "managers" at every level is a constant proportion of the aggregate salaries of the people they directly supervise. More precisely, if at every level i of the organization there are y_i employed, then the number of employed at the level below is $y_{i-1} = sy_i$. Furthermore, the

[132] See Section 4.4, pp. 20–24 in Piketty (2001a,b) and Appendix A in Roine et al. (2009).

[133] It is important to recall that in some countries the recent rise in top shares is not primarily driven by increased wage dispersion but by a return of capital in the top. This may be due partly to work-related remuneration taking the form of capital income for tax reasons, but also due to capital actually becoming more important. In this section, we will focus on theories that aim to explain increased concentration of earnings in the top.

wage w_i at any level i is related to the aggregate of those below by a fixed proportion, p, such that $w_i = nw_{i-1}p$. Under these assumptions the upper tail of the earnings distribution will be approximately Pareto distributed with exponent $\alpha = \frac{\log s}{\log(1+p)}$.[134] At any level in the organization the people above will earn on average a constant of the wage at that level, the multiple being $\alpha/(\alpha - 1)$.

How much pay increases as one moves up in such an organization is determined by the number of individuals supervised and by the pay increase at every level, but also by the size of the organization. If firms become larger in terms of the total number of employees, the salaries of the top management can be expected to increase. This basic insight, that large firms pay their top managers more than small firms, was noted by Mayer (1960) and is also a prominent fact in the data on the distribution of CEO pay (see, e.g., the overview by Murphy, 1999). But the hierarchical models have other problems empirically, in particular when it comes to explaining the very top. As noted already by Phelps Brown (1977, p. 309) plausible values of the span of control (the number of direct subordinates at each level in the hierarchy) and the pay raise at each step of the hierarchy do not match observed Pareto exponents (see Atkinson, 2008a, p. 77). Individuals high up in organizations simply earn more than what the model would predict.

In hierarchical models individuals are not paid based on "ability" but based on "responsibility."[135] But if "ability" determines the growth of a firm and the size of operations, then "responsibility" is endogenous, and the matching of ability and position becomes important.[136] In Rosen (1981) the distributions of firm size, span of control, and managerial incomes are modeled as the joint outcome of market assignments of personnel to hierarchical positions. Assuming the process assigns the most able individuals to the highest positions and that the talent of these individuals also multiplies throughout the organization, this results in firms of more capable managers being larger and also justifies high rewards to these managers.[137] In particular, it suggests that both the size distribution of firms and pay are skewed relative to the underlying ability distribution. Focusing on

[134] See Lydall (1968, pp. 127ff) and also Section 7.2.1.4 for more details.

[135] However, the relationship between the success of a firm, its growth, and consequently the size of the firm (both in sales and individuals employed) was in this way indirectly related to ability of the management, see Lydall (1968, Chapter 4). Also Baumol (1959, p. 46) at the time made the observation that "executive salaries appear to be far more closely correlated with the scale of the operations of the firm than with its profitability."

[136] This problem is the focus of many so-called assignment models, that in general study matching in perfectly competitive markets focusing on the combined effect of indivisibilities and heterogeneity on both sides of the market. In labor markets these features are important when there are complementarities between types of jobs and types of workers (e.g., Sattinger, 1979). See Sattinger (1993) for a review.

[137] The assumption that the process allocates the highest talent to the highest position is a contested one, and there is a large literature on executive compensation where many note that executive pay is not always based on performance, for example, Bertrand and Mullainathan (2001) and Bebchuk and Fried (2004).

CEO pay across different firms, Terviö (2008) built on this kind of assignment model for managerial talent to a distribution of firms, where firm size may be different not just due to managerial ability but for other reasons as well. Under the assumption that the larger firms will have most to benefit from hiring the best managers, the pay levels of these individuals across firms will be determined by distributions of firm size and managerial talent. In such a context the value of the highest talent may be significant for the largest firm. In similar spirit, Gabaix and Landier (2008) suggested that even very small ability differences can have large impacts on firm value. They found the sixfold increase of U.S. CEO pay between 1980 and 2003 can be fully attributed to the sixfold increase in market capitalization of large companies during that period.

A common feature in these (and many other) models is the idea that something (exogenous or endogenous) transforms small differences in the underlying ability to large differences in outcomes. Lazear and Rosen (1981) showed how compensation based on the outcome of a tournament where only the winner receives compensation can induce the highest effort under certain assumptions.[138] In general, attempts to implement payment schemes that give efforts to perform well has created a growth of performance-based pay in many fields. These schemes typically have the effect that the increase the individual returns to "top performers." However, it is not clear that the effect is positive for the economy as a whole or even for the implementing firm.[139] In models following Rosen (1981) a combination of technological change (production that makes replication easier such as printing, recording) and the size of the market gives the "most talented" disproportionately large rewards.[140] As the market reach of a so-called "superstar" increases, the returns to the highest talent also goes up, and at the same time the returns to those just below in the ability distribution goes down. The "global leader" drives out individuals or firms that used to be competitive at a more local level leading to increased concentration in top rewards. Frank and Cook (1995) argued that an increasing number of markets have developed features that fit the superstar model; they have become what they call "winner-take-all-markets." The examples range from activities

[138] This is, however, for example not true if the differences between the competitors is too large; see, for example, Freeman and Gelber (2010).

[139] See Fahlenbrach and Stulz (2009) and Bebchuck and Spamann (2010) for overviews of performance-based pay in the financial sector. Cahuc and Challe (2009) showed how performance pay can attract individual talent to a certain sector but also to possible misallocation of this talent in the economy. Agarwal and Wang (2009) showed how a shift to performance-based pay results in an increase in earnings dispersion but also to an increase in the amount of risk-taking with negative effects on aggregate performance.

[140] As Rosen noted in his paper, the basic idea was clear already to Marshall (1890/1920). Another early observer of the phenomenon, Watkins (1907) wrote: "The opportunity of the business man in any line to profit by value increase is multiplied by the increase in the breadth and in the number of exchanges" (pp. 62–63), and he went on to note that: "Even very slight changes in price, under modern conditions of a world-wide market and an unprecedented scale of individual transactions, may mean enormous gain or loss" (p. 63).

where broadcasting in a wide sense enlarges the market (such as markets for sports stars, artists, writers), to those where hiring a "superstar" may become more important as the amounts that hinges on their performance grows (lawyers, investment bankers, and CEOs), to more standard product markets where decreasing transportation and other trading costs make increases the potential market.[141]

So to what extent is the top of the distribution composed by such superstars? Kaplan and Rauh (2010) studied the representation of four sectors, top executives in nonfinancial firms, top employees in the financial sector (investment banks, hedge funds, and private equity), lawyers, and professional athletes and celebrities, in the top of the U.S. income distribution. They found that financial sector employees comprise a larger share than top executives from other sectors, and also that their share has grown in the past decades.[142] Athletes and celebrities as well as lawyers are certainly represented in the top but play a comparatively small role. Most striking perhaps is that the aggregate of these four groups account for less than 25% of the top income earners. This is due both to missing high-earning individuals in these four groups but also to the top of the income distribution consisting of much more than representatives of these groups.

Overall, theories focusing on various ways in which the underlying ability distribution may be magnified in terms of top earnings certainly contribute to our understanding of the recent increase in top income shares. There are also a number of areas where it seems clear these effects have grown over the past decades. But also some developments suggest that these theories are unlikely to be the full explanation, especially if one looks at the longer run developments. Frydman and Saks (2007) studied the ratio of CEO to worker pay in the United States over the period 1936–2005. They showed that this ratio was falling between the 1930s and the 1970s even though firms certainly grew in size over this period. Over this longer period they concluded that relationship between pay and firm growth is weak. In Figure 7.24, we complement their U.S. data with corresponding data from Sweden for the period 1950–2011. The long-run picture is very similar with falling ratios until around 1980 and then clear upturns thereafter. The level difference between the countries is marked, however, as is the fact that the recent increase has been much larger in the United States than in Sweden.

Another study that has looked at the long-run development of wages in a field with features that have been suggested to magnify small differences in ability, namely the financial sector, is Phillippon and Reshef (2012). They found that deregulation of financial markets is closely tied to compensation levels, as well as education levels and innovation, but also that the sector in the 1930s and since the 1990s seems to pay wages that are

[141] Gersbach and Schmutzler (2007) and Manasse and Turrini (2001) study how increased market size can drive increasing wage inequality within the top of the distribution.

[142] Bell and van Reenen (2010) also found that the financial sector is clearly overrepresented in the top of the U.K. income distribution.

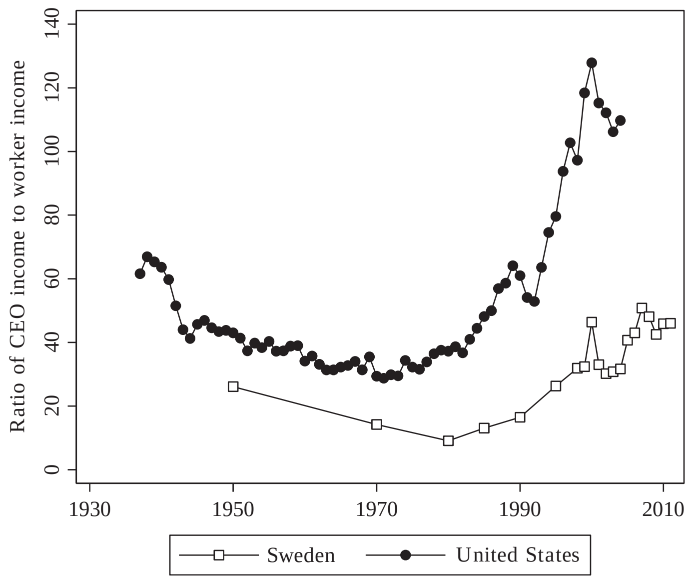


Figure 7.24 CEO and worker incomes in Sweden and the United States, 1936–2011. *Note and sources: Ratios are based on the following series. U.S. CEO incomes in 2005 U.S. dollars refer to CEOs in the largest 500 corporations in the ExecuComp database, from Frydman and Saks (2007) (including salary, bonus, long-term payments, and options granted). This series was generously shared by Carola Frydman. Average income refers to workers in the Social Security Administration database, collected from Kopczuk et al. (2007, table A2, downloaded at http://www.columbia.edu/~wk2110/uncovering/, 2013-12). Swedish incomes refer to CEOs in the 50 largest Swedish corporations, and male industrial workers, data coming from LO (2013, bilaga 2, "Näringslivet").*

substantially higher than what can be accounted for by observable factors (such as increased complexity of tasks and education levels). Interestingly, when comparing the relative pay in the financial sector with the top percentile income share in the entire United States, as is done in Figure 7.25, the resemblance is striking. The post-Depression drop in the 1930s is close to contemporaneous, and this is also true for the strong increase beginning in the late 1970s.

Finally, some scholars have pointed to the possibility of changing social norms as the most likely explanation for why top earnings have increased so much in recent decades (e.g., Levy and Temin, 2007; Piketty and Saez, 2003). Atkinson (2008a, Chapter 8) illustrates how, in a setting where individual utility depends on income as well as conforming to a social norm about fair pay (which operates both on the employer and employee sides of the market), there can be multiple equilibria for a given distribution of underlying ability.[143] The loss of utility when not adhering to the norm depends on how many others do the same. As a consequence, market forces alone do not uniquely determine the

[143] Similar to the model in Akerlof (1980).

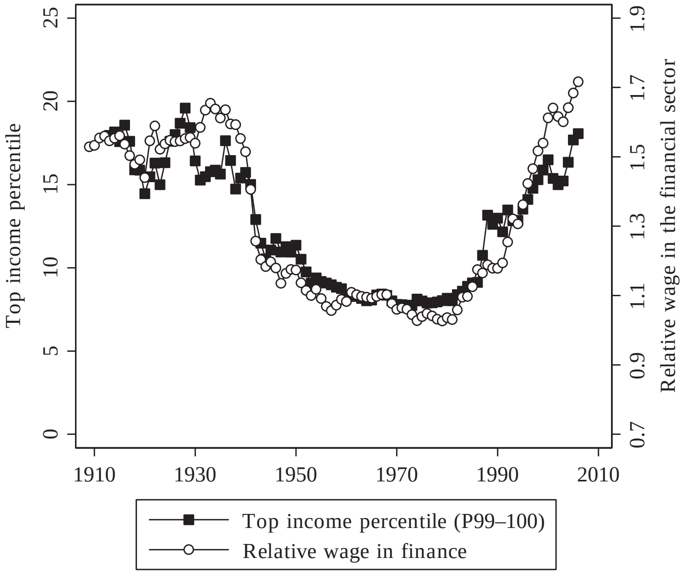


Figure 7.25 Relative wage in the financial sector versus the top income percentile in the United States, 1909–2006. *Note and sources: Ratio of financial sector wages to wages in agricultural and industrial sectors is from Phillippon and Reshef (2012) and U.S. top income share (excluding realized capital gains) are from Piketty and Saez (2003 and updates).*

outcome. There can be a situation where most individuals adhere to a norm, according to which pay is determined by a combination of ability and a fixed amount, as well as a situation where few individuals adhere to the norm, and pay is determined by individual productivity. Depending on initial conditions, different countries can converge on different pay norms, and "exogenous shocks" to the economy may cause a shift from one equilibrium to another.

7.4.4 Econometric Evidence on Determinants of Top Income Shares

A key objective in the top income project has been to create a sufficiently rich cross-country panel to enable an econometric testing of questions about what determines inequality.[144] In this subsection we will report on the results from a number of such studies.

7.4.4.1 Determinants of Inequality: Correlations over the Long Run

Roine et al. (2009) combined top income shares with data on a number of variables that have been suggested to affect inequality. The approach is not to test a particular theory

[144] As mentioned earlier, we focus on questions about what determines inequality, but obviously the top income data set has a large number of applications for questions regarding the consequences of inequality as well.

but rather to draw on a large number of models to produce a list of variables of interest in an exploratory fashion (see Atkinson and Brandolini, 2006, for a discussion of the clear limitations with such an approach). Econometrically the adopted method is to analyze first differences (using both first differenced generalized least squares and dynamic (with lagged dependent variable) first differences), assuming a linear relationship at least in this specification. Panel estimations make it possible to account for all unobservable time-invariant factors as well as common and country-specific trends. The potential that the relationships change over time is dealt with indirectly by allowing effects to differ over the level of development and for different country groups, and so on. This is clearly not the same as testing for the long-run effects of various variables on inequality but rather a way of testing what the short-run effects look like over the long run.

The main variables included in the analysis are the following. Financial development is measured as the sum of stock market capitalization and total amount of bank deposits divided by GDP. Trade openness is measured either *de facto* as the trade share in GDP (i.e., sum of exports and imports over GDP) or *de jure* as the average tariff rate (total tariffs paid divided by traded volume). Public sector influence is proxied by the share of central government spending in GDP and as the top marginal tax rates. Finally, we also include GDP per capita and population.[145] Given the importance of changes within the top, the income shares of three groups are analyzed: the rich (P99–100), the upper middle class (P90–99), and the rest of the population (P0–P90).[146]

Table 7.6 reports the regression results, and some basic relationships stand out as constantly robust across all specifications.[147] First, economic growth, that is, change in GDP per capita, seems to have been pro-rich over the twentieth century. In periods of faster than average growth, top income earners have benefited more than proportionally.[148] A likely reason for this result is simply that top incomes are (and have been) more closely related to performance than other incomes. This result is similar at different levels of

[145] There are also a number of additional variables, such as measures of democracy (Polity) and proxies for technological development (share of agriculture in GDP, number of patents) that are used in a robustness section. See Roine et al. (2009) for details.

[146] Clearly, any such division is arbitrary, but the results are not sensitive to the exact definitions of these top groups. Running the regressions defining the top 0.5% or the top 1.5% does not have any qualitative impact on the results. A threshold around top 1% can be justified by looking at the details of income compositions indicating that (approximately) the top 1% as a whole is very different from the rest of the top decile, especially with regard to capital income share. Also a similar classification, but with respect to wealth, is made in Hoffman et al. (2007).

[147] In Roine et al. (2009), the number of countries ranged between 12 and 14. Since then top income data has become available for more countries, and we have rerun the regressions. The results largely go through with the number of countries now ranging between 15 and 20. Here we also report results for some additional variables, such as Polity, that were not included in the results in Roine et al. (2009).

[148] See also Dew-Becker and Gordon (2005), who found that high productivity growth mainly benefitted the rich in the U.S. postwar era.

Table 7.6 Long-run determinants of top income shares

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	ΔTop1	**ΔTop1**	**ΔTop10–1**	**ΔTop10–1**	**ΔTop1**	**ΔTop1**	**ΔTop1**	**ΔTop1**
ΔGDPpc	4.44***	3.00***	–8.24***	–8.58***	4.28***	6.87***	2.76**	2.80**
	(0.850)	(1.150)	(1.421)	(1.652)	(0.827)	(1.674)	(1.164)	(1.186)
ΔPop	−8.75**	−6.07	−8.93	−7.30	−6.55*	−3.31	3.53	3.59
	(3.576)	(4.687)	(5.470)	(6.752)	(3.470)	(3.940)	(4.250)	(4.257)
ΔGovspend	−0.49	5.75	−7.68*	−9.97*	1.67	2.12	1.55	1.51
	(3.599)	(3.958)	(4.641)	(5.091)	(3.489)	(3.453)	(4.697)	(4.735)
ΔFindev	0.90***	0.94***	−0.25	−0.10	0.87***	0.79***		
	(0.197)	(0.295)	(0.214)	(0.318)	(0.232)	(0.177)		
ΔOpenness	−5.59***	−4.86**	1.51	4.88*	−4.78***	−5.15***	−0.42	−0.43
	(1.550)	(2.022)	(2.190)	(2.515)	(1.539)	(1.437)	(1.233)	(1.235)
ΔMarginaltax		−1.41*		−0.68				
		(0.758)		(0.761)				
ΔDemocracy					−0.82***	−0.54*		
					(0.278)	(0.325)		
ΔDemo.*GDPpc						−2.71*		
						(1.525)		
Bank crisis							−1.13***	−1.12***
							(0.426)	(0.431)
Currency crisis								0.03
								(0.502)
Observations	173	145	147	128	166	166	180	180
Countries	20	20	18	17	19	19	17	17

Notes: The regression method and underlying data are described in Section 7.4.4.1 and Roine et al. (2009). "Δ" denotes log change between 5-year periods, "GDPpc" is real per capita GDP, "Pop" is population, "Govspend" is central government spending as share of GDP, "Findev" is the sum of banking deposits and stock market capitalization as share of GDP, "Openness" is trade share in GDP, "Marginaltax" is the statutory top marginal income tax rate, "Democracy" is the autocracy score in the Polity IV index, and "Bank crisis" and "Currency crisis" are dummy variables for years when such crises occurred. Robust standard errors in parentheses. $^{*}p<0.1$, $^{**}p<0.05$, $^{***}p<0.01$.

development and is not different between Anglo-Saxon and other countries. However, the relation between economic growth and income share for the "upper middle class" (P90–99) seems to be the reverse. In high growth periods this group loses out in relative terms. This, again, highlights the importance of distinguishing between groups in the top decile. Second, financial development seems to have been pro-rich over the twentieth century, both in Anglo-Saxon countries and elsewhere and regardless of whether it is approximated using bank deposits or stock market capitalization (often said to be a difference between Continental Europe and Anglo-Saxon countries).[149] High marginal tax rates also have a consistent negative effect on top income, whereas government spending seems associated with a larger income share for the P0–P90 group. Somewhat surprisingly perhaps, there is little evidence of any clear effects of trade openness on top shares over the long run. Democracy, as measured by the commonly used Polity IV score, turns out to be negatively related to top shares, and this effect is larger as countries become richer, suggesting that democracy and development have an equalizing effect. The subsections that follow look at some of these relations in more detail.

7.4.4.2 The Effect of Top Tax Rates on Top Incomes

The theoretical effects of taxes on top incomes are not obvious. Recently much progress has been made in the field of optimal taxation both with respect to optimal labor income focusing on top income responses (Piketty et al., 2013) and in dynamic settings where income is determined by labor and capital and also influenced by bequests across generations (Piketty and Saez, 2013a,b).[150] Taken together the expected effect of higher top rates is to lower top income shares. This also appears to be the consistent finding in recent empirical work. Saez (2004) showed that changes in marginal tax rates over the period 1960–2000 can explain variation in top income shares in the United States, but also that the effect only seems to hold for the top percentile group. Saez and Veall (2005) showed that Canadian top income shares are negatively correlated with top marginal income tax rates. Using a similar specification, Roine and Waldenström (2008) concluded that changes in top rates in Sweden also had a significant effect on the Swedish development over the twentieth century, and Jäntti et al. (2010) concluded that the drop in top rates was a key determinant in the increase of Finnish top shares. Atkinson and Leigh (2013) found that top income shares are highly correlated across Anglo-Saxon countries and that top shares are very responsive to changes in marginal tax rates. Over the period 1970–2000 they estimated that reductions in tax rates could explain between one-third and one-half of the rise in the income share of the top percentile group. Atkinson and

[149] The finding that finance is pro-rich does not preclude that it can also be pro-poor, as has been found in previous research (e.g., Beck et al., 2007), but that it is the groups in the middle that seem to benefit the least from financial development.

[150] Persson and Sandmo (2005) however studied a tournament setting and showed that under some conditions increased taxation lead to increased inequality.

Leigh (2013) also tried to estimate the cumulative effect and found that a fall in the marginal tax rate on investment income (based on a lagged moving average) is associated with a rise in the share of the top 1%. Finally, Piketty et al. (2013) showed that there is a strong negative correlation between top tax rates and top 1% income shares in 18 OECD countries since 1960, and also that there is no evidence of high top shares corresponding to higher growth.

7.4.4.3 Political and Institutional Factors and the Impact of Crises

One potential advantage with the new top income data is that it spans a sufficiently long period for there to be sufficient variation in the degree of democracy and other institutional variables.[151] The long time period also makes it possible to potentially capture sufficiently many crises episodes to test effects of these econometrically.

The results in Table 7.6 include the role of democracy as captured by the well-known Polity IV measure. The results suggest that democracy indeed has an equalizing effect, but that it appears to be confined to reducing the top percentile's income share leaving the share of the rest of the top decile largely unchanged. Scheve and Stasavage (2009) looked at the effect of institutional differences in centralized wage bargaining and partisanship in 13 countries over the twentieth century but do not find any clear effects.

The effects of financial crises is also addressed in Table 7.6 and originally by Roine et al. (2009) using data from Bordo et al. (2001) and Laeven and Valencia (2008), indicating a negative effect of banking crises, but not currency crises, on the top percentile's income shares.[152] Atkinson and Morelli (2011), however, noted that when looking closer at how one might characterize the development after crises episodes, it is difficult to find a clear pattern. In many cases data are insufficient to give a clear picture of the direction of development.[153]

7.4.5 What Do We Learn?

The long-run development of wealth and income inequality is clearly a result of the joint effect of changes in both distributions. From any starting point, economic, social, and technological developments interact with shocks, crises, and policy to determine the evolution of both. The dynamics can go in both directions with effects over different time horizons. Exogenous events or policy can lead to wealth concentration going up or down, resulting in capital incomes becoming more or less important in explaining income inequality. But periods where high-income earners, due to exogenous factors or policy, receive a larger share of the total can also lead to increased wealth

[151] See Chapter 21 in this Handbook (Acemoglu et al., 2013) for an overview on the relation between democracy and inequality.

[152] However, currency crises and banking crises do not seem to have any clear effects.

[153] Bordo and Meissner (2012) used top income data to study if inequality has been related to credit expansion which in turn has been argued to be a good predictor of crises. They found no evidence of such a relationship.

concentration. Over time this leads to a return of capital incomes, and unless individuals consume all their earnings over their lifetime, inheritance also becomes a factor across generations.

In terms of understanding inequality developments over the past century in the countries studied in this chapter, some main themes are worth recapitulating. The drop in inequality over the first half of the century is mainly due to decreased wealth shares of top wealth holders, resulting in declining capital incomes. The wage share of high earners, however, typically looks very stable. The drop in wealth holdings and subsequently in capital incomes seems to be the result of both macroshocks such as the World Wars and financial crises but also to policies pursued in many countries. After these shocks high top marginal tax rates made it difficult to rapidly accumulate new fortunes, and inequality leveled out or continued to decline. Such a development can be accounted for in a simple model that combines capital and earnings and uses it to study the effects of exogenous shocks to the capital stock and the effects of taxation.

The recent increase in inequality, observable in many countries but not all, seems to be mainly due to increased top wages. Explaining this turns the focus to a different set of explanations emphasizing higher returns on the labor market for some groups (based on higher ability, skill, effort, education, etc.). Two key facts seem important in guiding efforts to understand this change. First, much of the increase is concentrated to a small fraction at the top of the population. This means that theories focusing on changes for broader groups (such as "skilled" and "unskilled") at least need to be complemented by a mechanism explaining the increase within the top. Second, the degree to which top earnings have increased relative to the average is very different across countries. Thus, a theory based on a common global shift of some kind at least needs to be complemented with mechanisms that can account for the cross-country difference.

Finally, the preliminary econometric evidence points to taxation being important in explaining the developments. Even though magnitudes in the short run may seem small, it is important to take the long-run dynamic effects into account. Financial development and economic growth being pro-rich also stand out as clear and robust correlates over the whole of the twentieth century, but so far we have only begun to use the data for systematic cross-country analysis.

7.5. SUMMARY AND CONCLUDING REMARKS

In this chapter we have outlined the broad facts about long-run trends in the distribution of income and wealth. The focus has been on findings primarily stemming from the top income literature and recent studies of wealth concentration, using historical tax and estate data. However, we have also tried to relate the new results to previous observations in the economic growth and economic history literatures. The end result is always going to be subjective, and we have therefore tried to be as clear as possible on where in the development there is disagreement.

When it comes to *describing* the overall developments of income inequality across the 26 countries studied in Section 7.2, there are three possible broad eras that can usefully be distinguished. The first is the period before the First World War, the second is the period from around 1914 until 1980, and the third consists of the time thereafter. In the first period evidence is relatively clear on the fact that inequality was historically high in the beginning of the twentieth century. To what extent this high level was present throughout the nineteenth century or if it gradually increased is, however, still less clear due to the lack of data. There are some signs of increased inequality but many studies also point toward high and relatively stable levels before the twentieth century.

The period from around 1914 to 1980 is characterized by substantial drops in top income shares in almost all countries for which we have data. The top percentile share falls from around 20% before 1914 to between 5% and 10% around 1980. The decreasing income share for the lower parts of the top decile group are much more modest. In fact, in some countries the income share of lower half of the top decile group (P90–95) remains almost constant throughout the twentieth century. Thus, distinguishing developments within the top group seem important. Large parts of the decreases seem to happen in connection to shocks such as the World Wars or the Great Depression, but it is worth noting that decreases also take place in countries that did not take part in the war, such as Sweden. Also the drop continues after the Second World War throughout the high-growth periods in the 1950s and 1960s. In terms of income composition, most of the drop seems related to decreasing capital income.

The development after 1980 is less homogenous. In some countries, especially the United States and the United Kingdom, inequality has risen sharply. This increase has taken place from a level that was already high in relation to others before it started. In countries like Sweden and Finland, increases have also been substantial but here from internationally low levels to levels that are much higher but remain among the lowest. In, for example, France, Germany, and Japan, there is no clear upward trend but in absolute terms inequality remains higher than in the Nordic countries.

Turning to the development of the wealth distribution in the 10 countries for which we have long-run data, studied in Section 7.3, a picture similar to that of income emerges. In most places (the United States being the notable exception) wealth concentration was relatively constant and historically high before the twentieth century. Even if cross-country comparisons should be made with caution, there seems to have been important level differences. Estimated top wealth shares at the beginning of the twentieth century were clearly higher in the United Kingdom and in France than, for example, in the United States and in Switzerland, Finland, and Norway, with Denmark, Sweden, and the Netherlands in between. Starting around the First World War, the top percentile group wealth shares decreased substantially until around 1980. Thereafter the development is again more diverse but also much more uncertain and debated than for income

In terms of *understanding* these developments, we have, in Section 7.4, discussed a number of suggested theories and empirical regularities that aim at explaining various aspects. The developments over the first half of the twentieth century points to the importance of understanding the joint developments of wealth and income as much of the decrease is related to sharp drops in capital incomes in the top. The cumulative impact of taxation over time also seems important, especially for understanding the lack of recovery of top income shares in the decades after the Second World War. With respect to the different developments after 1980, it seems likely that many factors interact. There are probably important changes in terms of technological change and globalization that affect inequality, but the differences across countries also suggest that the impact depends on individual country characteristics (such as the functioning of the labor market, the education systems, and other policies). In most countries much of the inequality increase is driven by changes *within* the top, suggesting that an explanation must include a mechanism that gives an increasing income advantage to the very top groups, rather than only accounting for differences between broad groups such as, for example, skilled and unskilled. In addition, the recent increase is not homogenous in terms of income composition. In some countries (the United States) an increased earnings dispersion explains most of the recent increase, whereas in other countries (Sweden) capital seems more important. Finally, we also note that there are cases where data and explanations seem to fit what we observe since the 1980s, but when one applies a long-run perspective the same theory seems less successful (an example being executive pay as explained by the growth of firms). This need not imply that the account for the post-1980 period is incorrect, but it does suggest that most explanations are likely too sensitive to interaction with aspects that change both over time and across countries.

7.5.1 Going Forward

When looking ahead a number of areas seem promising in terms of future research. First, the ongoing work of extending the top income database is obviously important, both in terms increasing the number of countries, but also in terms of adding new dimensions. In at least some countries it could, for example, be possible to distinguish income for men and women over much of the twentieth century. Constructing similar data sets on long-run wealth inequality trends would also represent important contributions.

Second, making use of the top income database seems important. Numerous aspects of long-run developments can now be studied over a time span previously void of systematic inequality data. In doing so it is, as emphasized by Atkinson and Brandolini (2006), it is important to take an integrated approach to theory and estimation and to use proper econometric techniques to address deficiencies in the data.

Third, results on the importance of changes within the top illustrate how development can be missed or misinterpreted if one focuses solely on overall inequality. These findings also pose challenges for theories trying to explain the recent surge in income inequality. It is in general quite likely that different explanations apply to different parts of the distribution.

Fourth, a number of the recent findings illustrate the importance of the interplay between wealth and work in determining total income. Also, when income is determined not just by actions over an individual's lifetime, inheritance also becomes important. This has important effects on a wide range of issues such as optimal taxation (Piketty and Saez, 2013b) and the future development of inheritance flows (Piketty, 2011).

These are some of the research areas that will surely be important for our understanding of long-run inequality trends, their causes, and their consequences.

ACKNOWLEDGMENTS

We have received valuable comments from Anthony Atkinson and seminar participants at the conference "Recent Advances in the Economics of Income Distribution" held at the Paris School of Economics, April 2013. We thank Wiemer Salverda for sharing wealth distribution data on the Netherlands. Data underlying all graphs and tables in the chapter are available on the authors' Web sites or on request.

APPENDIX

Table 7.A1 Sources of the historical wealth inequality data

Country	Year(s)	Wealth holder unit	Source
Australia	1915, 1987, 2002, 2006, 2010	Households	Katic and Leigh (2013, table A2)
	1953–1979	Adults	Katic and Leigh (2013, table A1)
Denmark	1789	Males > 19 years	Soltow (1985: table 4)
	1908–1975	Households	Zeuthen (1928: table IV 4: 521) for 1908–1925, Bjerke (1956: table 32) for 1939–1945, Statistics Denmark, *Statistisk Årbog* for 1950–1975. See also Alvaredo et al. (2013) using roughly the same sources
	1995–1996	Households	Statistics Denmark (1995, 1996: table 2)
Finland	1800	Males > 19 years	Soltow (1980, table 3)
	1922	Households	Soltow (1980, table 3)
	1926	Households	Soltow (1980, table 3)
	1967	Households	Soltow (1980, table 3)
	1987–2005	Adults	Statistics Finland, tax statistics
France	1807–2010	Adults	Piketty et al. (2004: table A4, 2006), Piketty (2014)
Netherlands	1894–1974	Adults	Wilterdink (1984)
	1993–2011	Households (survey)	Statistics Netherlands (2010), Salverda et al. (2013). Series are submitted by Wiemer Salverda
Norway	1789	Households	Soltow (1980: table 3)
	1868	Households (?)	Mohn (1873: 10, 30)
	1912	Households	Statistics Norway (1915a: 6*, 20*–21*)
	1930	Households	Statistics Norway (1934: 63*f)
	1948–2011	Households (1983–1993 are adjusted individuals as described in text)	Statistics Norway, *Statistisk Årbok* and Statistikdatabasen (see text)
	2009	Households	Epland and Kirkeberg (2012, table 8)

Table 7.A1 Sources of the historical wealth inequality data—cont'd

Country	Year(s)	Wealth holder unit	Source
Sweden*	1800	Males > 19 years	Soltow (1985: tables 4, 5)
	1908	Households	Finansdepartementet (1910: 31)
	1920	Households	Statistics Sweden (1927)
	1930	Households	Statistics Sweden (1937, 1938)
	1935	Households	Statistics Sweden (1940)
	1937	Households	SOU (1942: 52)
	1945	Households	Statistics Sweden (1951)
	1946–1950	Households	SOS Skattetaxeringarna
	1951	Households	Statistics Sweden (1956)
	1966	Households	SOU (1969: 54)
	1970	Households	SOS Inkomst och Förmögenhet 1970, Budgetundersökningen
	1975	Households	Spånt (1979)
	1978–1998	Households	Jansson and Johansson (2000, table 15)
	1999–2007	Households	Own calculations based tabulated household distributions retrieved from Statistics Sweden's Wealth Register (see Roine and Waldenström, 2009, for details)
	1873–1877	Individuals	Finansdepartementet (1879)
	1906–1908	Individuals	Finansdepartementet (1910). For 1908 there is also wealth data based on applying the estate multiplier method (Finansdepartementet, 1910: 14–34)
	1954	Individuals	SOU (1957). See Roine and Waldenström (2009) for details
	1967	Individuals	SOU (1969). See Roine and Waldenström (2009) for details
	2002–2003	Individuals	SOU (2004). See Roine and Waldenström (2009) for details
Switzerland	1913–1997	Households	Dell et al. (2007: table 3)
United Kingdom (England and Wales before 1938)	1740, 1810, 1875	Adults	Lindert (2000: table 2)
	1911–1913	Adults	Atkinson and Harrison (1978: table 6.1)
	1923–1977	Adults	Atkinson et al. (1989a,b: table 1)
	1978–2005	Adults	Inland Revenue Statistics (2006: table 13.5)
United States	1774	Adults > 19 years	Shammas (1993: table 2)
	1916–2000	Adults > 19 years	Kopczuk and Saez (2004: table 3)
	1774	Households (free adult men and unmarried women)	Lindert (2000, table 3)
	1860	Households (free adult male heads of households)	Shammas (1993: table 2)
	1890	Families	Lindert (2000: table 3)
	1922–1958	Households	Wolff (1996: table 1).
	1962–2010	Households	Kennickell (2009, table 4; 2011, table 1)

Note: List of sources and data definitions of the wealth distribution data used in the chapter. See text for further descriptions of the data. The definition of household used here is not exactly identical across (and sometimes even within) country samples. The basic concept is one where individuals (aged 18 or above) and married couples count as one household (see Section 7.3.2.1.7 on Sweden for details). The asterisk denotes the way that pages are numbered/indexed in the original publications.

Table 7.A2 Top wealth percentile (P99–100) share of total private wealth in 10 countries

Year	Australia	Denmark	Finland	France	Netherlands	Norway	Sweden	Switzerland	United Kingdom	USA (hh)	USA (ind)
1740									43.6		
1774										16.5	28.0
1789		56.0				47.0					
1800			34.0				48.0				
1810				45.6					54.9		
1820				46.7							
1830				47.5							
1840				46.0							
1850				50.3							
1860				52.0						21.0	
1868						36.0					
1870				50.4							
1875									61.1		
1880				49.5							
1890				51.1						25.8	
1894					54.0						
1900				58.7							
1905					55.0						
1908		46.3					53.8				
1909			37.4								
1910				60.5							
1911									69.0		
1912						37.2					
1914					56.5						
1915	33.9	47.0	35.9					42.3			
1916											35.6
1917		44.1									35.6
1918		43.6									36.8
1919		42.6			50.0			36.4			39.9
1920		37.2		49.2			51.5				37.6
1921		39.7						38.1			35.2

1922		39.6	25.4							36.7	36.0
1923		39.9							60.9		35.2
1924		39.3							59.9		36.7
1925		38.7			47.5			40.7	61.0		36.0
1926			29.7						57.3		35.1
1927									59.8		39.2
1928									57.0		36.5
1929								42.0	55.5	44.2	36.8
1930				47.4	48.0	37.6	50.0		57.9		40.3
1931											34.7
1932											28.4
1933										33.3	30.3
1934								40.4			28.1
1935					42.0		42.8				27.8
1936								40.1	54.2		29.7
1937							42.7				27.0
1938								44.4	55.0		27.1
1939		41.7			45.0					36.4	26.0
1940				36.3				40.4			25.3
1941								41.5			25.3
1942											23.7
1943											24.3
1944		39.2									25.5
1945							37.7	37.1		29.8	24.7
1946							37.7				24.5
1947							34.7	38.3			24.3
1948						34.6	34.1				23.0
1949		31.3					33.2	37.8		27.1	22.6
1950		29.6		33.4			32.8		47.2		22.8
1951		29.7			34.0		32.2	39.0	45.8		
1952		29.4							43.0		

Continued

Table 7.A2 Top wealth percentile (P99–100) share of total private wealth in 10 countries—cont'd

Year	Australia	Denmark	Finland	France	Netherlands	Norway	Sweden	Switzerland	United Kingdom	USA (hh)	USA (ind)
1953	14.6	29.5						40.0	43.6	31.2	23.8
1954	10.7	29.3							45.3		23.2
1955	9.9	29.5			35.0			41.5	44.5		
1956	12.4	27.1							44.5		24.7
1957	9.4	27.2						41.9	43.4		
1958	9.6	27.1							41.4		24.2
1959	9.6	27.9							41.4		
1960	9.4	26.4		31.9	37.5	25.5			33.9		25.2
1961	9.0	26.7							36.5		
1962	7.8	26.9							31.4	31.8	24.4
1963	9.0	27.2									
1964	8.8	27.6							34.5		
1965	8.9	24.2			33.0				33.0		24.7
1966	7.8	24.8					23.4		30.6		
1967	6.9	24.6	19.4						31.4		
1968	6.3								33.6		
1969	9.0							41.6	31.1	31.1	22.9
1970	8.9	24.8		22.0	31.0		20.1		29.7		
1971	7.9	25.5							28.4		
1972	10.0	25.3							31.7		23.1
1973	6.8					21.5			27.3		
1974	8.9				28.0				22.6		
1975	7.8	25.9					17.0		22.7		
1976	6.8					19.5			24.4		19.3
1977	7.4								22.1		
1978	6.6						16.6		20.0		
1979						18.5			20.0		
1980				22.0					19.0		
1981								33.0	18.0		

1982						18.0			18.0		19.1
1983						17.5	17.7		20.0	33.8	21.1
1984						18.0			18.0		21.0
1985						18.9	16.5		18.0		22.4
1986						18.7			18.0		22.7
1987	9.7		16.1			18.7			18.0		21.6
1988						18.9	18.4		17.0		21.7
1989			15.8			18.9			17.0	30.1	22.0
1990			15.3	21.7		18.8	20.7		18.0		20.9
1991			15.2			18.8		33.6	17.0		21.5
1992			15.3			17.5	19.5		18.0	30.2	21.2
1993			13.7		22.1	17.0			18.0		21.3
1994			13.9		22.5	17.7			19.0		21.6
1995		26.9	13.9		20.7	18.0			19.0	34.6	21.5
1996		27.2	15.1		23.0	18.3			20.0		21.4
1997			16.2		22.2	19.1	20.3	34.8	22.0		21.2
1998			17.3		23.6	18.7			22.0	33.9	21.7
1999			21.2		22.2	19.2	19.3		23.0		21.7
2000			21.2	23.5	22.2	19.6	21.9		23.0		20.8
2001			20.6			18.5	19.7		22.0	32.7	
2002	11.8		19.6			18.0	18.0		24.0		
2003			20.4			18.3	17.9		21.0		
2004			20.7			19.2	20.5			33.4	
2005			21.6			20.4	19.7				
2006	16				21.4	21.0	18.5				
2007					22.0	22.1	18.8			33.8	
2008					21.5	21.9					
2009			22.7		19.8	21.2					
2010	11.4			24.4	22.3	19.4				34.5	
2011					23.0	18.9					

Note: Note that many series contain several breaks in data definitions that may severely affect comparability both over time and across countries. See Table 7.A1 for sources and Section 7.3 for details. "USA (hh)" denotes top wealth shares in the U.S. distribution of households (with some exception) and "USA (ind)" top wealth shares in the U.S. distribution of adult individuals.

Table 7.A3 Bottom four percentiles in top 5 wealth percentiles (P95–99) as share of total private wealth in nine countries

Year	Denmark	Finland	France	Netherlands	Norway	Sweden	Switzerland	United Kingdom	USA
1740								30.0	
1774									24.5
1789	24.0				23.0				
1800		25.5				21.4			
1807			33.1						
1810								19.4	
1817			31.5						
1827			35.0						
1837			35.2						
1847			33.4						
1857			29.5						
1860									28.0
1867			29.4						
1875								13.0	
1877			27.2						
1887			27.9						
1902			24.3						
1908	31.7					22.4			
1909		21.2							
1911								18.0	
1912					32.0				
1913							26.9		
1914									
1915	27.2	20.9		24.0			26.4		
1917	27.9								
1918	26.7								
1919	26.9						25.9		
1920	29.0					27.7			
1921	27.8			26.0			25.9		
1922	29.0	23.4							
1923	28.3							21.1	
1924	28.2							21.6	
1925	29.4			25.0			23.9	21.1	
1926		23.4						22.6	
1927								21.5	
1928								22.6	
1929							24.6	23.4	
1930				26.0	33.0	27.3		21.3	
1934							27.5		
1935				26.0		28.0			
1936							28.0	23.2	
1938							28.9	22.2	

Table 7.A3 Bottom four percentiles in top 5 wealth percentiles (P95–99) as share of total private wealth in nine countries—cont'd

Year	Denmark	Finland	France	Netherlands	Norway	Sweden	Switzerland	United Kingdom	USA
1939	28.8			26.0					
1940							27.2		
1941							27.9		
1944	29.2								
1945						28.3	27.2		
1946						28.1			
1947			26.3			28.7	27.1		
1948					27.8	29.0			
1949	26.9					28.7	27.2		
1950	26.4			26.0		27.8		27.2	
1951	26.3					26.8	27.3	27.9	
1952	26.3							27.4	
1953	26.2						26.6	27.7	
1954	26.1			26.0				26.7	
1955	25.7						25.8	27.0	
1956	24.6							27.1	
1957	24.6						25.5	25.7	
1958	24.8							26.8	
1959	24.7							26.1	
1960	23.9			27.0	25.5			25.6	
1961	23.9							24.3	
1962	23.8							23.5	21.3
1963	23.6								
1964	23.3							24.5	
1965	22.0							25.4	
1966	22.3			26.0		23.5		25.1	
1967	22.4	27.3						24.9	
1968								25.0	
1969							25.2	25.3	17.7
1970	22.9			25.0		22.0		24.2	
1971	23.2							24.2	
1972	22.7							25.2	
1973					22.5			24.2	
1974								26.0	
1975	24.6			26.0		21.0		23.8	
1976					22.8			24.6	
1977								24.3	
1978						22		17.0	
1979					23.6			17.0	
1980								17.0	
1981							23.6	18.0	

Continued

Table 7.A3 Bottom four percentiles in top 5 wealth percentiles (P95–99) as share of total private wealth in nine countries—cont'd

Year	Denmark	Finland	France	Netherlands	Norway	Sweden	Switzerland	United Kingdom	USA
1982					27.5			18.0	
1983					26.9	20.7		17.0	22.8
1984					24.6			17.0	
1985					24.1	20.5		18.0	
1986					24.3			18.0	
1987		20.4			24.7			19.0	
1988					24.3	21.7		19.0	
1989		20.1			24.0			18.0	24.1
1990		19.6			24.2	21.9		17.0	
1991		19.4			23.9		23.0	18.0	
1992		19.3			22.8	21.4		20.0	24.4
1993		17.9		23.7	17.6			20.0	
1994		18.0	26.0	22.8	17.8			20.0	
1995	27.0	18.0		22.4	17.5			19.0	21.3
1996	25.8	18.3		22.2	17.6			20.0	
1997		18.6		21.8	17.6	23.8	23.2	21.0	
1998		18.7		22.1	17.1			18.0	23.3
1999		19.0		21.9	17.4	24.7		20.0	
2000		18.9		21.4	17.5	22.5		21.0	
2001		17.9			17.4	22.4		20.0	25.0
2002		17.9			17.5	23.2		21.0	
2003		17.9			17.7	22.8		19.0	
2004		17.8			16.5	22.2			24.1
2005		17.7			16.9	23.9		19.0	
2006				21.9	17.2	21.5			
2007				20.9	17.7	21.8			26.6
2008				21.0	17.7				
2009		18.6		22.2	18.0				
2010				22.4	18.3				28.4
2011				22.2	17.7				

Note: Note that many series contain several breaks in data definitions that may severely affect comparability both over time and across countries. For the United States, the series is based on estimates from mainly the household distribution. See Table 7.A1 for sources and Section 7.3 for details.

Table 7.A4 Top wealth decile (P90–100) as share of total private wealth in eight countries

	Denmark	Finland	France	Norway	Sweden	Switzerland	United Kingdom	USA
1740							86.0	
1774								59.0
1789	88.0			81.0				
1800		75.8			86.0			
1810			79.9				83.4	
1820			81.8					
1830			83.2					
1840			80.4					
1850			82.4					
1860			83.7					
1870			81.8					
1875							83.8	
1880			84.6					
1890			84.7					72.2
1908	87.3				86.0			
1909		70.6						
1910			88.5					
1911							92.0	
1912				76.3				
1913						84.8		
1915	84.8	69.1				80.5		
1917	85.9							
1918	85.4							
1919	83.4					76.3		
1920	80.9		81.7		91.7			
1921	83.6					77.0		
1922	83.9	64.4						
1923	83.6						89.1	
1924	83.2						88.1	
1925	83.7					75.8	88.4	
1926		68.3					87.4	
1927							88.3	
1928							87.2	
1929						76.7	86.3	
1930			80.0	84.6	89.5		86.6	
1935					83.6			
1936							85.7	
1938							85.0	
1939	84.0							
1940			75.8			80.8		
1941						81.9		
1944	82.4							

Continued

Table 7.A4 Top wealth decile (P90–100) as share of total private wealth in eight countries—cont'd

	Denmark	Finland	France	Norway	Sweden	Switzerland	United Kingdom	USA
1945					83.2	78.3		
1946					81.4			
1947					79.6	79.0		
1948				78.4	80.7			
1949	73.2				79.1	78.8		
1950	71.1		72.8		77.3			
1951	71.1				75.0	79.9		
1952	70.7							
1953	70.7					79.9		
1954	70.5							
1955	70.1					79.9		
1956	66.8							
1957	66.9					79.9		
1958	66.8							
1959	67.4							
1960	65.2		69.9	66.4			71.5	
1961	65.4						71.7	
1962	65.4						67.3	64.6
1963	65.1							
1964	65.1						71.4	
1965	60.9						71.7	
1966	61.9				63.2		69.2	
1967	61.8	61.9					70.0	
1968							71.6	
1969						78.9	67.7	
1970	62.9		62.0		57.9		68.7	
1971	63.9						67.6	
1972	62.8						70.4	
1973				58.7				
1975	67.5				54.0			
1976				57.1			50.0	
1977							50.0	
1978					54.5		49.0	
1979				58.1			50.0	
1980			61.8				50.0	
1981						69.6	50.0	
1982				58.6			49.0	
1983				56.5	54.5		50.0	68.9
1984				57.1			48.0	
1985				57.4	53.4		49.0	
1986				56.7			50.0	
1987		50.7		56.9			51.0	
1988				56.4	56.6		49.0	
1989		50.6		55.7			48.0	67.2

Table 7.A4 Top wealth decile (P90–100) as share of total private wealth in eight countries—cont'd

	Denmark	Finland	France	Norway	Sweden	Switzerland	United Kingdom	USA
1990		50.0	61.0	55.9	58.7		47.0	
1991		49.5		55.5		69.9	47.0	
1992		49.1		53.0	57.7		50.0	67.1
1993		45.9		47.7			51.0	
1994		46.2		48.6			52.0	
1995	73.3	46.2		48.7			50.0	67.8
1996	72.2	47.6		48.9			52.0	
1997		49.0		49.5	61.1	71.3	54.0	
1998		50.0		51.7			52.0	68.6
1999		52.9		52.5	60.3		55.0	
2000		52.8	62.1	52.8	59.9		56.0	
2001		51.4		51.6	57.7		54.0	69.8
2002		50.5		51.2	57.3		54.0	
2003		51.1		51.9	56.6		53.0	
2004		51.4		52.9	57.8			69.5
2005		51.9		54.2	58.4		54.0	
2006				54.6	55.9			
2007				55.7	56.7			71.5
2008				55.3				
2009		54.4		52.6				
2010			62.4	51.1				74.5
2011				49.5				

Note: Note that many series contain several breaks in data definitions that may severely affect comparability both over time and across countries. For the United States, the series is based on estimates from mainly the household distribution. See Table 7.A1 for sources and Section 7.3 for details.

REFERENCES

Aaberge, R., Atkinson, A.B., 2010. Top incomes in Norway. In: Atkinson, A.B., Piketty, T. (Eds.), Top Incomes over the Twentieth Century: A Global Perspective. Oxford University Press, Oxford.

Acemoglu, D., Autor, D.H., 2012. What does human capital do? A review of Goldin and Katz's race between education and technology. J. Econ. Lit. 50 (2), 426–463.

Acemoglu, D., Autor, D.H., 2013. Skills, tasks and technologies: implications for employment and earnings. In: Ashenfelter, O., Card, D.E. (Eds.), Handbook of Labor Economics, vol. 4. North-Holland, Amsterdam.

Acemoglu, D., Johnson, S., Robinson, J.A., 2005. Institutions as the fundamental cause of long-run growth. In: Aghion, P., Durlauf, S. (Eds.), Handbook of Economic Growth. North-Holland, Amsterdam.

Acemoglu, D., Naidu, S., Restrepo, P., Robinson, J.A., 2013. Democracy, redistribution and inequality. In: Atkinson, A.B., Bourguinon, F. (Eds.), Handbook of Income Distribution. In: vol. 2. North-Holland, Amsterdam.

Adelman, I., Taft Morris, C., 1973. Economic Growth and Social Equity in Developing Countries. Stanford University Press, Stanford.

Agarwal, S., Wang, F.H., 2009. Perverse incentives at the banks? Evidence from a natural experiment. Working Paper Series WP-09-08, Federal Reserve Bank of Chicago.

Aghion, P., Caroli, E., Garcia-Penalosa, C., 1999. Inequality and economic growth: the perspective of the new growth theories. J. Econ. Lit. 37 (4), 1615–1660.

Akerlof, G.A., 1980. A theory of social custom, of which unemployment may be one consequence. Q. J. Econ. 94 (4), 749–775.
Alvaredo, F., 2010. The rich in Argentina over the twentieth century: 1932–2004. In: Atkinson, A.B., Piketty, T. (Eds.), Top Incomes: A Global Perspective, vol. 2. Oxford University Press, Oxford.
Alvaredo, F., Atkinson, A.B., 2011. Colonial rule, apartheid and natural resources: top incomes in South Africa, 1903–2007. CEPR Working Paper No. 8155, Centre for Economic Policy Research, London.
Alvaredo, F., Gasparini, L., 2014. Post-1970 trends in inequality and poverty in developing and emerging countries. In: Atkinson, A.B., Bourguinon, F. (Eds.), Handbook of Income Distribution. In: vol. 2. North-Holland, Amsterdam.
Alvaredo, F., Pisano, E., 2010. Top incomes in Italy, 1974–2004. In: Atkinson, A.B., Piketty, T. (Eds.), Top Incomes: A Global Perspective, vol. 2. Oxford University Press, Oxford.
Alvaredo, F., Saez, E., 2009. Income and wealth concentration in Spain from a historical and fiscal perspective. J. Eur. Econ. Assoc. 7 (5), 1140–1167.
Alvaredo, F., Vélez, L.J., 2013. High incomes and personal taxation in a developing economy: Colombia 1993–2010: CEQ Working Paper No. 12, Commitment to Equity.
Alvaredo, F., Sandholt Jensen, P., Sharp, P., 2013. Inheritance and Wealth in Denmark. Mimeo.
Andreoni, J., Erard, B., Feinstein, J., 1998. Tax compliance. J. Econ. Lit. 36 (2), 818–860.
Armour, P., Burkhauser, R.V., Larrimore, J., 2013. Deconstructing income inequality measures: a crosswalk from market income to comprehensive income. Am. Econ. Rev. Papers Proc. 103 (3), 173–177.
Atkinson, A.B., 1997. Bringing income distribution in from the cold. Econ. J. 107 (441), 297–321.
Atkinson, A.B., 1999. Is Rising Income Inequality Inevitable? A Critique of the Transatlantic Consensus WIDER Annual Lecture, vol 3. UNU-WIDER.
Atkinson, A.B., 2005. Top incomes in the UK over the 20th century. J. Roy. Stat. Soc. A Stat. Soc. 168 (2), 325–343.
Atkinson, A.B., 2007a. Methodological issues. In: Atkinson, A.B., Piketty, T. (Eds.), Top Incomes over the Twentieth Century: A Contrast Between European and English-Speaking Countries. Oxford University Press, Oxford.
Atkinson, A.B., 2007b. The distribution of top incomes in the United Kingdom, 1908–2000. In: Atkinson, A.B., Piketty, T. (Eds.), Top Incomes over the Twentieth Century: A Contrast Between Continental European and English-Speaking Countries. Oxford University Press, Oxford.
Atkinson, A.B., 2008a. The Changing Distribution of Earnings in OECD Countries Oxford and New York. Oxford University Press, Oxford.
Atkinson, A.B., 2008b. Concentration among the rich. In: Davies, J.B. (Ed.), Personal Wealth from a Global Perspective. Oxford University Press, Oxford.
Atkinson, A.B., 2010. Top incomes in a rapidly growing economy: Singapore. In: Atkinson, A.B., Piketty, T. (Eds.), Top Incomes: A Global Perspective, vol. 2. Oxford University Press, Oxford.
Atkinson, A.B., 2011. Income Distribution and Taxation in Mauritius: A Seventy-five Year History of Top Incomes. Mimeo.
Atkinson, A.B., Brandolini, A., 2001. Promise and pitfalls in the use of 'Secondary' data-sets: income inequality in OECD countries as a case study. J. Econ. Lit. 39, 771–799.
Atkinson, A.B., Brandolini, A., 2006. The panel-of-countries approach to explaining income inequality: an interdisciplinary research agenda. In: Morgan, S.L., Grusky, D.B., Fields, G.S. (Eds.), Mobility and Inequality: Frontiers of Research in Sociology and Economics. Stanford University Press, Stanford.
Atkinson, A.B., Harrison, A., 1978. The Distribution of Personal Wealth in Britain. Cambridge University Press, Cambridge.
Atkinson, A.B., Leigh, A., 2007a. The distribution of top incomes in Australia. In: Atkinson, A.B., Piketty, T. (Eds.), Top Incomes over the Twentieth Century: A Contrast Between European and English-Speaking Countries. Oxford University Press, Oxford.
Atkinson, A.B., Leigh, A., 2007b. The distribution of top incomes in New Zealand. In: Atkinson, A.B., Piketty, T. (Eds.), Top Incomes over the Twentieth Century: A Contrast Between European and English-Speaking Countries. Oxford University Press, Oxford.

Atkinson, A.B., Leigh, A., 2013. The distribution of top incomes in five Anglo-Saxon countries over the long run. Econ. Rec. 89 (Suppl. S1), 31–47.

Atkinson, A. B., Morelli, S., 2011. Economic Crises and Inequality. Human Development Reports Research Paper 2011/06. UNDP.

Atkinson, A.B., Piketty, T. (Eds.), 2007. Top Incomes over the Twentieth Century: A Contrast Between European and English-Speaking Countries. Oxford University Press, Oxford.

Atkinson, A.B., Piketty, T. (Eds.), 2010. Top Incomes: A Global Perspective. Oxford University Press, Oxford.

Atkinson, A.B., Salverda, W., 2005. Top incomes in the Netherlands and the United Kingdom over the 20th century. J. Eur. Econ. Assoc. 3 (4), 883–913.

Atkinson, A.B., Søgaard, J., 2013. The long-run history of income inequality in Denmark: top incomes from 1870 to 2010, EPRU Working Paper 2013-01.

Atkinson, A.B., Gordon, J.P.F., Harrison, A., 1989. Trends in the shares of top wealth-holders in Britain: 1923–1989. Oxf. Bull. Econ. Stat. 51 (3), 315–332.

Atkinson, A.B., Piketty, T., Saez, E., 2010. Top Incomes in the Long Run of History. In: Atkinson, A.B., Piketty, T. (Eds.), Top Incomes: A Global Perspective, vol. II. Oxford University Press, Oxford.

Atkinson, A.B., Piketty, T., Saez, E., 2011. Top incomes in the long run of history. J. Econ. Lit. 49 (1), 3–71.

Auerbach, A.J., 1989. Capital gains taxation and tax reform. Natl. Tax J. 42, 391–401.

Autor, D., Katz, L.F., Kearney, M., 2006. The polarization of the U.S. labor market. Am. Econ. Rev. Papers Proc. 96 (2), 189–194.

Banerjee, A., Piketty, T., 2009. Top Indian incomes, 1922–2000. In: Atkinson, A.B., Piketty, T. (Eds.), Top Incomes: A Global Perspective. In: vol. 2. Oxford University Press, Oxford.

Baumol, W., 1959. Business Behavior, Value and Growth. Macmillan, New York, 1959.

Bebchuck, L.A., Spamann, H., 2010. Regulating bankers' pay. Georgetown Law J. 98 (2), 247–287.

Bebchuk, L.A., Fried, J., 2004. Pay Without Performance: The Unfulfilled Promise of Executive Compensation. Harvard University Press, Cambridge.

Beck, T., Demirgüç-Kunt, A., Levine, R., 2007. Finance, inequality and the poor. J. Econ. Growth 12 (1), 27–49.

Bell, B., van Reenen, J., 2010. Bankers' pay and extreme wage inequality in the UK. CEP Special Report.

Bengtsson, T., van Poppel, F., 2011. Socioeconomic inequalities in death from past to present: an introduction. Explor. Econ. Hist. 48, 343–356.

Bengtsson, N., Holmlund, B., Waldenström, D., 2012. Lifetime versus annual tax progressivity: Sweden, 1968–2009. UCFS Working Paper, Uppsala University.

Benhabib, J., Bisin, A., 2007. The distribution of wealth and redistributive policies, Meeting Papers 368, Society for Economic Dynamics.

Bentzel, R., 1953. Inkomstfördelningen i Sverige. Almqvist & Wiksell, Uppsala.

Bentzen, J., Schmidt-Sørensen, J.B., 1994. Wealth distribution and mobility in Denmark: a longitudinal study. CLS Working Paper 4, Aarhus School of Business, Aarhus.

Berg, L., 1983. Konsumtion och sparande—en studie av hushållens beteende. Ph.D. Dissertation, Department of Economics, Uppsala University.

Bertrand, M., Mullainathan, S., 2001. Are CEOs rewarded for luck? The ones without principals are. Q. J. Econ. 116 (3), 901–932.

Besley, T., Persson, T., 2009. The origins of state capacity: property rights, taxation, and policy. Am. Econ. Rev. 99, 1218–1244.

Besley, T., Persson, T., 2013. Taxation and development. In: Auerbach, A., Chetty, R., Feldstein, M., Saez, E. (Eds.), Handbook of Public Economics. North-Holland, Amsterdam.

Bjerke, K., 1956. Changes in Danish income distribution 1939–52. Income Wealth IV, 98–154.

Björklund, A., 1998. Income distribution in Sweden: what is the achievement of the welfare state? Swed. Econ. Policy Rev. 5, 39–80.

Björklund, A., Jäntti, M., Lindquist, M.J., 2009. Family background and income during the rise of the welfare state: brother correlations in income for Swedish men born 1932–1968. J. Public Econ. 93 (5–6), 671–680.

Björklund, A., Roine, J., Waldenström, D., 2012. Intergenerational top income mobility in Sweden: capitalist dynasties in the land of equal opportunity? J. Publ. Econ. 96 (5–6), 474–484.
Bohlin, J., Lundh, C., Prado, S., 2011. Markets or politics? The Swedish urban–rural wage gap, 1865–1945. Unpublished manuscript, Gothenburg University.
Bordo, M.D., Meissner, C.H., 2012. Does inequality lead to a financial crisis?. NBER Working Paper 17896, National Bureau of Economic Research.
Bordo, M.D., Eichengreen, B., Klingebiel, D., Soledad Martinez-Peria, M., 2001. Is the crisis problem growing more severe? Econ. Policy 16 (32), 51–82.
Borgerhoff Mulder, M., Bowles, S., Hertz, T., Bell, A., Beise, J., Clark, G., Fazzio, I., Gurven, M., Hill, K., Hooper, P., Irons, W., Kaplan, H., Leonetti, D.L., Marlowe, F., Naidu, S., Nolin, D., Piraino, P., Quinlan, R.J., Sear, R., Shenk, P., Smith, E.A., Wiessner, P., 2009. The intergenerational transmission of wealth and the dynamics of inequality in pre-modern societies. Science 326, 682–688.
Bourguignon, F., 1981. Pareto superiority of unegalitarian equilibria in Stiglitz' model of wealth distribution with convex saving function. Econometrica 49 (6), 1469–1475.
Bowley, A.L., 1914. The British super-tax and the distribution of income. Q. J. Econ. 28 (2), 255–268.
Brenner, Y., Kaelble, H., Thomas, M. (Eds.), 1991. Income Distribution in Historical Perspective. Cambridge University Press, Cambridge.
Burkhauser, R.V., Feng, S., Jenkins, S., Larrimore, J., 2012. Recent trends in top income shares in the USA: reconciling estimates from March CPS and IRS tax return data. Rev. Econ. Stat. 94 (2), 371–388.
Burkhauser, R.V., Hahn, M., Wilkins, R., 2013. Measuring top incomes using tax record data: a cautionary tale from Australia. NBER Working Paper No. 19121.
Cahuc, P., Challe, E., 2009. Produce or speculate? Asset bubbles, occupational choice and efficiency. IZA Discussion Paper 4630.
Caselli, F., 1999. Technological revolutions. Am. Econ. Rev. 89 (1), 78–102.
Christensen, H.M., 2003. Skatteberegningsreglerne gennem 100 år. Skatteministeriet, Copenhagen.
Clark, G., 2008. A Farewell to Alms. A Brief History of the World. Princeton University Press, Princeton, NJ.
Coghlan, T., 1906. Discussion. J. Roy. Stat. Soc. 69, 735–736.
Comin, D., Mestieri, M., 2013. If technology has arrived everywhere, why has income diverged?. NBER Working Papers 19010, National Bureau of Economic Research.
Cowell, F., 2011. Measuring inequality. Oxford University Press, Oxford.
Cowell, F., 2013. UK wealth inequality in an international context. In: Hills, J. (Ed.), Wealth in the UK. Oxford University Press, Oxford.
Crum, W.L., 1935. Individual shares in the national income. Rev. Econ. Stat. 17 (6), 116–130.
Davies, J.B., Shorrocks, A.F., 2000. The distribution of wealth. In: Atkinson, A.B., Bourguignon, F. (Eds.), Handbook of Income Distribution, vol. 1. North-Holland, Amsterdam.
De Long, J.B., 1996. Billionaires. www.j-bradford-delong.net/Econ_Articles/billionaires.html (accessed 10.07.06).
Dell, F., 2007. Top incomes in Germany throughout the twentieth century. In: Atkinson, A.B., Piketty, T. (Eds.), Top Incomes over the Twentieth Century: A Contrast Between European and English-Speaking Countries. Oxford University Press, Oxford.
Dell, F., Piketty, T., Saez, E., 2007. Income and wealth concentration in Switzerland over the twentieth century. In: Atkinson, A.B., Piketty, T. (Eds.), Top Incomes over the Twentieth Century: A Contrast Between European and English-Speaking Countries. Oxford University Press, Oxford.
Dew-Becker, I., Gordon, R.J., 2005. Where did the productivity growth go? Inflation dynamics and the distribution of income. Brooking Papers Econ. Act. 2, 67–127.
Edlund, L., Kopczuk, W., 2009. Women, wealth and mobility. Am. Econ. Rev. 99 (1), 146–178.
Eloranta, J., García-Iglesias, C., Jalava, J., Ojala, J., 2006. On the road to prosperity: an introduction. In: Ojala, J., Eloranta, J., Jalava, J. (Eds.), On the Road to Prosperity. An Economic History of Finland. SKS, Helsinki.
Epland, J., Kirkeberg, M.I., 2012. Wealth distribution in Norway: evidence from a new register-based data. Statistics Norway, Report 35/2012.

Esping-Andersen, G., 1990. The Three Worlds of Welfare Capitalism. Princeton University Press, Princeton, NJ.

Fahlenbrach, R., Stulz, R., 2009. Bank CEO incentives and the credit crisis. J. Financ. Econ. 99 (1), 11–26.

Feenberg, D.R., Poterba, J.M., 1993. Income inequality and the incomes of very high-income taxpayers: evidence from tax returns. In: Poterba, J.M. (Ed.), Tax Policy and the Economy, vol. 7. MIT Press, Cambridge and London, pp. 145–177.

Feldstein, M., 1976. Social security and the distribution of wealth. J. Am. Stat. Assoc. 71, 800–807.

Finansdepartementet, 1879. Sammandrag öfver stämpelafgifterna för lagfarter, inteckningar, äktenskapsförord, morgongåfvobref och afhandlingar om lösöresköp under år 1877 samt för bouppteckningar under åren 1873–1877. Finansdepartementet, Stockholm.

Finansdepartementet, 1910. Bouppteckningar efter aflidna, inregistrerade vid vederbörande domstolar åren 1906–1908. Finansdepartementet, Stockholm.

Frank, R., Cook, P.J., 1995. The Winner-Take-All Society: How More and More Americans Compete for Ever Fewer and Bigger Prizes, Encouraging Economic Waste, Income Inequality, and an Impoverished Cultural Life. Free Press, New York.

Frankel, S.H., Herzfeld, H., 1943. European income distribution in the Union of South Africa and the effect thereon of income taxation. S. Afr. J. Econ. 11 (2), 121–136.

Freeman, R., 2011. Globalization and inequality. In: Salverda, W., Nolan, B., Smeeding, T.M. (Eds.), The Oxford Handbook on Economic Inequality. Oxford University Press, Oxford.

Freeman, R., Gelber, A.M., 2010. Prize structure and information in tournaments: experimental evidence. Am. Econ. J. 2 (1), 149–164.

Frick, J., Grabka, M., 2013. Public pension entitlement and the distribution of wealth. In: Gornick, J.C., Jäntti, M. (Eds.), Income Inequality. Economic Disparities and the Middle Class in Affluent Countries. Stanford University Press, Stanford.

Friedman, E., Johnson, S., Kaufman, D., Zoido-Lobaton, P., 2000. Dodging the grabbing hand: the determinants of unofficial activity in 69 countries. J. Public Econ. 76 (3), 459–493.

Friesen, S.J., Scheidel, W., 2009. The Size of the Economy and the Distribution of Income in the Roman Empire.

Frydman, C., Saks, R.E., 2007. Executive compensation: a new view from a long-term perspective, 1936–2005. Rev. Financ. Stud. 23 (5), 2099–2138.

Gabaix, X., Landier, A., 2008. Why has CEO pay increased so much? Q. J. Econ. 123 (1), 49–100.

Gale, W.G., 1998. The effects of pensions on household wealth: a reevaluation of theory and evidence. J. Polit. Econ. 106 (4), 706–723.

Gersbach, H., Schmutzler, A., 2007. Does globalization create superstars?. CEPR Discussion Papers 6222, Centre for Economic Policy Research, London.

Goldin, C., Katz, L., 1998. The origins of technology-skill complementarity. Q. J. Econ. 113 (3), 693–732.

Goldin, C., Katz, L.F., 2008. The Race Between Education and Technology. The Belknap Press of Harvard University Press, Cambridge, MA.

Grossman, G.M., Rossi-Hansberg, E., 2008. Trading tasks: a simple theory of offshoring. Am. Econ. Rev. 98 (5), 1978–1997.

Haig, R.M., 1921. The Federal Income Tax. Columbia University Press, New York.

Hoffman, P.T., Postel-Vinay, G., Rosenthal, J.-L., 2007. Surviving Large Losses: Financial Crises, the Middle Class, and the Development of Capital Markets. Harvard University Press, Cambridge, MA.

Inland Revenue Statistics, 2006. Distribution among the adult population of marketable wealth. http://www.hmrc.gov.uk/stats/personal_wealth/table13_5.xls (accessed 22.06.06).

Jansson, K., Johansson, L., 2000. Förmögenhetsfördelningen i Sverige 1997 med tillbakablick till 1975. Statistics Sweden, Stockholm.

Jenkins, S., Brandolini, A., Micklewright, J., Nolan, B., 2013. The Great Recession and the Distribution of Household Income. Oxford University Press, Oxford.

Jäntti, M., 2006. Trends in the distribution of income and wealth: Finland 1987–1998. In: Wolff, E.N. (Ed.), International Perspectives on Household Wealth. Edward Elgar, Northampton.

Jäntti, M., Riihela, M., Sullstrom, R., Tuomala, M., 2010. Trends in top income shares in Finland. In: Atkinson, A.B., Piketty, T. (Eds.), Top Incomes: A Global Perspective, vol. 2. Oxford University Press, Oxford.

Johannessen, N., Zucman, G., 2014. The end of bank secrecy? an evaluation of the G20 tax haven crack-down. Am. Econ. J. Econ. Pol. 6 (1), 65–69.

Johansson, F., Klevmarken, A.N., 2007. Comparing survey and register wealth data. In: Essays on Measurement Error and Nonresponse. Uppsala University Economic Studies 103, Uppsala University, Uppsala.

Johnson, N.O., 1935. The brookings report on inequality in income distribution. Q. J. Econ. 49 (4), 718–724.

Johnson, N.O., 1937. The pareto law. Rev. Econ. Stat. 19, 20–26.

Jones, A.H., 1970. Wealth estimates for the American middle colonies, 1774. Econ. Dev. Cult. Change (4), Part 2, pp. i-x + 1–172.

Jones, A.H., 1972. Wealth estimates for the New England colonies about 1770. J. Econ. Hist. 32, 98–127.

Jones, A.H., 1980. Wealth of a Nation to Be: The American Colonies on the Eve of the American Revolution. Columbia University Press, New York.

Journal of Roman Studies, 99, 61–91.

Jutikkala, E., 1953. The distribution of wealth in Finland in 1800. Scand. Econ. Hist. Rev. 1 (1), 81–103.

Kanbur, R., 2000. Income distribution and development. In: Atkinson, A.B., Bourguignon, F. (Eds.), Handbook of Income Distribution, vol. 1. North-Holland, Amsterdam.

Kaplan, S., Rauh, J., 2010. Wall street and main street: what contributes to the rise in the highest incomes? Rev. Financ. Stud. 23 (3), 1004–1050.

Katic, P., Leigh, A., 2013. Top wealth shares in Australia 1915–2012. Working Paper, International Water Management Institute.

Katz, L., Autor, D., 1999. Changes in the wage structure and earnings inequality. In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics, vol. 3. North-Holland, Amsterdam.

Katz, L., Murphy, K., 1992. Changes in relative wages: supply and demand factors. Q. J. Econ. 107, 35–78.

Kennickell, A., 2009. Ponds and Streams: Wealth and Income in the U.S., 1989 to 2007. Finance and Economics Discussion Series 2009–2013, Board of Governors of the Federal Reserve System (U.S.), Washington, D.C.

Kennickell, A., 2011. Tossed and Turned: Wealth Dynamics of U.S. Households 2007–2009. Finance and Economics Discussion Series 2011–51, Board of Governors of the Federal Reserve System (U.S.), Washington, D.C.

Kiær, A.N., 1917. Indtaegts- og formuesforhold efter skatteligningerne for 1913/14 sammenignet med tidligere og senere aar. Departementet for Sociale Saker, Kristiania.

Kopczuk, W., Saez, E., 2004. Top wealth shares in the United States, 1916–2000: evidence from estate tax returns. Natl. Tax J. 57 (2), 445–487.

Kopczuk, W., Saez, E., Song, J., 2007. The American Dream: Inequality and Mobility in Social Security Earnings Data Since 1937, NBER Working Paper No. 13345. National Bureau of Economic Research.

Kopczuk, W., Saez, E., Song, J., 2010. Earnings inequality and mobility in the United States: evidence from social security data since 1937. Q. J. Econ. 125 (1), 91–128.

Kuznets, S., 1953. Shares of Upper Income Groups in Income and Savings. National Bureau of Economic Research, Cambridge, Mass.

Kuznets, S., 1955. Economic growth and income inequality. Am. Econ. Rev. 45 (1), 1–28.

Kuznets, S., 1963. Quantitative aspects of the economic growth of nations: VIII. Distribution of income by size. Econ. Dev. Cult. Change 11 (2, part 2), 1–80.

Laeven, L., Valencia, F., 2008. Systemic banking crises: a new database. IMF Working Paper No. 08/224.

Lampman, R.J., 1962. The Share of Top Wealth-Holders in National Wealth: 1922–56. Princeton University Press, Princeton, NJ.

Lane, P.R., Milesi-Feretti, G.M., 2001. The external wealth of nations: measures of foreign assets and liabilities for industrial and developing countries. J. Int. Econ. 55, 263–294.

Lane, P.R., Milesi-Feretti, G.M., 2007. The external wealth of nations mark II. J. Int. Econ. 73, 223–250.

Lavindkomstkommissionen, 1979. Udviklingen i formuefordelingen 1960–1977. Delrapport 2, Lavindkomstkommissionen, København.
Lazear, E.P., Rosen, S., 1981. Rank-order tournaments as optimum labor contracts. J. Polit. Econ. 89, 841–864.
Leamer, E.E., 2007. A flat world, a level playing field, a small world after all, or none of the above? A review of Thomas L Friedman's the World is flat. J. Econ. Lit. 45 (1), 83–126.
Leigh, A., 2005. Deriving long-run inequality series from tax data. Econ. Rec. 81 (255), 58–70.
Leigh, A., 2007. How closely do top income shares track other measures of inequality? Econ. J. 117 (524), F619–F633.
Leigh, A., 2009. Top incomes. In: Salverda, W., Nolan, B., Smeeding, T. (Eds.), The Oxford Handbook of Economic Inequality. Oxford University Press, Oxford.
Leigh, A., van der Eng, P., 2009. Top incomes in Indonesia: 1920–2004. In: Atkinson, A.B., Piketty, T. (Eds.), Top Incomes: A Global Perspective, vol. 2. Oxford University Press, Oxford.
Leigh, A., Jencks, C., Smeeding, T.M., 2011. Health and economic inequality. In: Nolan, B., Salverda, W., Smeeding, T.M. (Eds.), The Oxford Handbook of Economic Inequality. Oxford University Press, Oxford.
Levy, P., Temin, P., 2007. Inequality and institutions in 20th century America. NBER Working Paper No. 13106.
Lindahl, M., Palme, M., Sandgren-Massih, S., Sjögren, A., 2012. Dynastic capital and long-term intergenerational persistence of earnings and educational attainment. Working Paper, Uppsala University.
Lindert, P.H., 1986. Unequal English wealth since 1670. J. Polit. Econ. 94 (6), 1127–1162.
Lindert, P.H., 2000. Three centuries of inequality in Britain and America. In: Atkinson, A.B., Bourguignon, F. (Eds.), Handbook of Income Distribution, vol. 1. North-Holland, Amsterdam.
Lindert, P., Williamson, J. G., 2003. Does globalization make the world more unequal? In: Bordo, M., Taylor, A., Williamson, J.G. (Eds.), Globalization in Historical Perspective. University of Chicago Press and National Bureau of Economic Research.
Lindert, P.H., Williamson, J.G., 2014. American incomes since the seventeenth century: gaining world leadership and losing equality. Princeton University Press, Princeton, NJ.
Ljungberg, J., 2006. The secular decline of professional relative earnings in Sweden. Evidence and explanation. Unpublished manuscript, Lund University.
LO, 2013. Makteliten—klyftorna består. En studie av inkomstutvecklingen för makteliten perioden 1950 till 2011. Landsorganisationen, Stockholm.
Long, J., Ferrie, J., 2007. The path to convergence: intergenerational occupational mobility in Britain and the US in three eras. Econ. J. 117, C61–C71.
Lydall, H., 1968. The Structure of Earnings. Oxford University Press, London.
Mallet, B., 1908. A method of estimating capital wealth from the estate duty statistics. J. Roy. Stat. Soc. 71, 65–84.
Manasse, P., Turrini, A., 2001. Trade, wages, and 'superstars'. J. Int. Econ. 54 (1), 97–117.
Marshall, A., 1920. Principles of Economics, eighth ed. Macmillan, London.
Mayer, T., 1960. Distribution of ability and earnings. Rev. Econ. Stat. 42 (2), 189–195.
Meade, J.E., 1964. Efficiency, Equality, and the Ownership of Property. Allen and Unwin, London.
Melitz, M.J., 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. Econometrica 71 (6), 1695–1725.
Melitz, M.J., Ottaviano, G., 2008. Market size, trade, and productivity. Rev. Econ. Stud. 75, 295–316.
Milanovic, B., 2006. An estimate of average income and inequality in Byzantium around year 1000. Rev. Income Wealth 52 (3), 449–470.
Milanovic, B., 2013. The Inequality Possibility Frontier: The Extensions and New Applications. Mimeo, World Bank.
Milanovic, B., Lindert, P., Williamson, J.G., 2011. Pre-industrial inequality. Econ. J. 121, 255–272.
Moehling, C.M., Steckel, R., 2001. Rising inequality: trends in the distribution of wealth in industrializing New England. J. Econ. Hist. 61 (1), 160–183.
Mohn, J.R., 1873. Statistiske bidrag til belysning af privatformuens fordeling i Norge. Norsk Retstidende 1–2, 1–32.

Mokyr, J., 1990. The Lever of Riches: Technological Creativity and Economic Progress. Oxford University Press, Oxford.
Moriguchi, C., Saez, E., 2008. The evolution of income concentration in Japan, 1885–2002: evidence from income tax statistics. Rev. Econ. Stat. 90, 713–734.
Morrisson, C., 2000. Historical perspectives on income distribution: the case of Europe. In: Atkinson, A.B., Bourguignon, F. (Eds.), Handbook of Income Distribution, vol 1. Elsevier, North-Holland, Amsterdam, pp. 217–260.
Murphy, K.J., 1999. Executive compensation. In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics, vol. 3. North-Holland, Amsterdam.
Nolan, B., 2007. Long term trends in top income shares in Ireland. In: Atkinson, A.B., Piketty, T. (Eds.), Top Incomes over the Twentieth Century: A Contrast Between European and English-Speaking Countries. Oxford University Press, Oxford.
Office for National Statistics, 2011. The Effects of Taxes and Benefits on Household Income, 2009/10. Further Analysis and Methodology. Office for National Statistics, London, UK.
Ohlsson, H., Roine, J., Waldenström, D., 2008. Long-run changes in the concentration of wealth: an overview of recent findings. In: Davies, J.B. (Ed.), Personal Wealth from a Global Perspective. Oxford University Press, Oxford.
O'Rourke, K.H., Williamson, J.G., 1999. Globalization and History. The Evolution of the 19th Century Atlantic Economy. MIT Press, Cambridge, MA.
O'Rourke, K.H., Taylor, A., Williamson, J.G., 1996. Factor price convergence in the late nineteenth century. Int. Econ. Rev. 37, 499–530.
Pareto, V., 1897. Cours d'economie politique, vol. ii, F. Rouge, Lausanne.
Paukert, F., 1973. Income distribution at different levels of development: a survey of evidence. Int. Lab. Rev. 108, 97–125.
Persson, M., Sandmo, A., 2005. Taxation and tournaments. J. Public Econ. Theory 7 (4), 543–559.
Phelps Brown, H., 1977. The Inequality of Pay. University of California Press, Berkeley.
Phillippon, T., Reshef, A., 2012. Wages and human capital in the U.S. financial industry: 1909–2006. Q. J. Econ. 127 (4), 1551–1609.
Piketty, T., 2000. Theories of persistent inequality and intergenerational mobility. In: Atkinson, A.B., Bourguignon, F. (Eds.), Handbook of Income Distribution, vol. 1. North-Holland, Amsterdam.
Piketty, T., 2001a. Les hauts revenus en France au 20ème siècle. Grasset, Paris.
Piketty, T., 2001b. Income inequality in France, 1900–1998. CEPR Discussion Paper No. 2876, Centre for Economic Policy Research, London.
Piketty, T., 2003. Income inequality in France, 1901–1998. J. Polit. Econ. 111 (5), 1004–1042.
Piketty, T., 2005. Top income shares in the long run: an overview. J. Eur. Econ. Assoc. 3 (2–3), 1–11.
Piketty, T., 2006. The Kuznets' curve, yesterday and tomorrow. In: Banerjee, A., Benabou, R., Mookerhee, D. (Eds.), Understanding Poverty. Oxford University Press, Oxford.
Piketty, T., 2007. Income inequality in France, 1900–1998. In: Atkinson, A.B., Piketty, T. (Eds.), Top Incomes over the Twentieth Century: A Contrast Between European and English-Speaking Countries. Oxford University Press, Oxford.
Piketty, T., 2011. On the long-run evolution of inheritance: France 1820–2050. Q. J. Econ. 126 (3), 1071–1131.
Piketty, T., 2014. Capital in the 21st Century. Harvard University Press, Cambridge, MA.
Piketty, T., Qian, N., 2009. Income inequality and progressive income taxation in China and India, 1986–2015. Am. Econ. J. Appl. Econ. 1 (2), 53–63.
Piketty, T., Saez, E., 2003. Income inequality in the United States, 1913–1998. Q. J. Econ. 118 (1), 1–39.
Piketty, T., Saez, E., 2006. The evolution of top incomes: a historical and international perspective. Am. Econ. Rev. Papers Proc. 96 (2), 200–205.
Piketty, T., Saez, E., 2007. How progressive is the U.S. federal tax system? A historical and international perspective. J. Econ. Perspect. 21 (1), 3–24.
Piketty, T., Saez, E., 2013a. Optimal labor income taxation. In: Handbook of Public Economics. vol. 5. North-Holland, Amsterdam.

Piketty, T., Saez, E., 2013b. A theory of optimal inheritance taxation. Econometrica 81 (5), 1851–1886.
Piketty, T., Zucman, G., 2013. Capital is back: wealth-income ratios in rich countries, 1700–2010. Working PaperParis, School of Economics.
Piketty, T., Zucman, G., 2014. Capital is back: wealth-income ratios in rich countries, 1700–2010. Q. J. Econ. 129 (3), 1255–1310.
Piketty, T., Postel-Vinay, G., Rosenthal, J.-L., 2004. Wealth concentration in a developing economy: Paris and France, 1807–1994. CEPR Working Paper 4631Centre for Economic Policy Research, London.
Piketty, T., Postel-Vinay, G., Rosenthal, J.-L., 2006. Wealth concentration in a developing economy: Paris and France, 1807–1994. Am. Econ. Rev. 96 (1), 236–256.
Piketty, T., Saez, E., Stantcheva, S., 2013. Optimal taxation of top incomes: a tale of three elasticities. Am. Econ. J. Econ. Pol. 6 (1), 230–271.
Poterba, J.M., Feenberg, D.R., 2000. The income and tax share of very high-income households, 1960–1995. Am. Econ. Rev. 90 (2), 264–270.
Reich, R., 2010. Aftershock. The Next Economy and America's Future. Alfred A. Knopf.
Roine, J., Waldenström, D., 2008. The evolution of top incomes in an egalitarian society: Sweden, 1903–2004. J. Public Econ. 92, 366–387.
Roine, J., Waldenström, D., 2009. Wealth concentration over the path of development: Sweden, 1873–2006. Scand. J. Econ. 111 (1), 151–187.
Roine, J., Waldenström, D., 2010. Top incomes in Sweden over the twentieth century. In: Atkinson, A.B., Piketty, T. (Eds.), Top Incomes: A Global Perspective, vol. 2. Oxford University Press, Oxford.
Roine, J., Waldenström, D., 2011. Common trends and shocks to top incomes: a structural breaks approach. Rev. Econ. Stat. 93 (3), 832–846.
Roine, J., Waldenström, D., 2012. On the role of capital gains in Swedish income inequality. Rev. Income Wealth 58 (3), 569–587.
Roine, J., Vlachos, J., Waldenström, D., 2009. The long-run determinants of inequality: what can we learn from top income data? J. Public Econ. 93 (7–8), 974–988.
Rosen, S., 1981. The economics of superstars. Am. Econ. Rev. 71 (5), 845–858.
Rosenbloom, J.L., Stutes, G.W., 2008. Reexamining the distribution of wealth in 1870. In: Rosenbloom, J. L. (Ed.), Quantitative Economic History: The Good of Counting. Routledge, London.
Saez, E., 2004. Reported incomes and marginal tax rates, 1960–2000: evidence and policy implications. In: Poterba, J. (Ed.), Tax Policy and the Economy. In: vol. 18. MIT Press, Cambridge, MA.
Saez, E., Veall, M.R., 2005. The evolution of high incomes in Northern America: lessons from Canadian evidence. Am. Econ. Rev. 95, 831–849.
Saez, E., Slemrod, J., Giertz, E., 2012. The elasticity of taxable income with respect to marginal tax rates: a critical review. J. Econ. Lit. 50, 3–50.
Salverda, W., 7, A.B., 2007. Top incomes in the Netherlands over the twentieth century. In: Atkinson, A.B., Piketty, T. (Eds.), Top Incomes Over the Twentieth Century: A Contrast Between Continental European and English-Speaking Countries. Oxford University Press, Oxford.
Salverda, W., Haas, C., de Graaf-Zijl, M., Lancee, B., Notten, N., Ooms, T., 2013. Country Report for the Netherlands: Growing Inequalities and Their Impacts in the Netherlands. GINI Project, Amsterdam.
Sattinger, M., 1979. Differential rents and the distribution of earnings. Oxford Econ. Papers 31, 60–71.
Sattinger, M., 1993. Assignment models of the distribution of earnings. J. Econ. Lit. 31, 831–880.
Scheve, K.F., Stasavage, D., 2009. Institutions, partisanship, and inequality in the long run. World Polit. 61 (2), 215–253.
Scheve, K.F., Stasavage, D., 2010. The conscription of wealth: mass warfare and the demand for progressive taxation. Int. Org. 64 (3), 529–561.
Scheve, K.F., Stasavage, D., 2012. Democracy, war, and wealth: evidence from two centuries of inheritance taxation. Am. Polit. Sci. Rev. 106 (1), 82–102.
Seldon, A., 1979. Tax Avoision: The Economic, Legal and Moral Inter-relationships Between Avoidance and Evasion. Institute of Public Affairs, London.

Shammas, C., 1993. A new look at long-term trends in wealth inequality in the United States. Am. Hist. Rev. 98 (4), 412–432.
Simon, H.A., 1957. The compensation of executives. Sociometry 20 (I), 32–35.
Simons, H., 1938. Personal Income Taxation: The Definition of Income as a Problem of Fiscal Policy. University of Chicago Press, Chicago.
Slemrod, J., 1992. Why People Pay Taxes: Tax Compliance and Enforcement. University of Michigan Press, Ann Arbor.
Slemrod, J., 1996. High-income families and the tax changes of the 1980s: the anatomy of behavioral response. In: Feldstein, M., Poterba, J. (Eds.), Empirical Foundations of Household Taxation. University of Chicago Press and the National Bureau of Economic Research, pp. 169–188.
Slemrod, J., 2000. The economics of taxing the rich. In: Slemrod, J. (Ed.), Does Atlas Shrug? The Economic Consequences of Taxing the Rich. Russell Sage Foundation and Harvard University Press.
Slemrod, J., Yitzhaki, S., 2002. Tax avoidance, evasion, and administration. In: Auerbach, A.J., Feldstein, M. (Eds.), Handbook of Public Economics. In: vol. 3. North-Holland, Amsterdam.
Smeeding, T., Thompson, J., Morelli, S., 2014. Post-1970 trends in within-country inequality and poverty. In: Atkinson, A.B., Bourguignon, F. (Eds.), Handbook of Income Distribution. In: vol. 2. North-Holland, Amsterdam (Chapter 9).
Smith, J.D., 1984. Trends in the concentration of personal wealth in the United States, 1958 to 1976. Rev. Income Wealth 30 (4), 419–428.
Söderberg, J., 1991. Wage differentials in Sweden, 1725–1950. In: Brenner, Y.S., Kaelble, H., Mark, T. (Eds.), Income Distribution in Historical Perspective. Cambridge University Press, Cambridge.
Soltow, L., 1968. Long-run changes in British income inequality. Econ. Hist. Rev. 21 (1), 17–29.
Soltow, L., 1971. Economic inequality in the United States in the period from 1790 to 1860. J. Econ. Hist. 31, 822–839.
Soltow, L., 1980. Wealth distribution in Norway and Denmark in 1789. Hist. Tidsskr. 59, 221–235.
Soltow, L., 1981a. Wealth distribution in Finland in 1800. Scand. Econ. Hist. Rev. 27 (1), 21–32.
Soltow, L., 1981b. Wealth distribution in Denmark in 1789. Scand. Econ. Hist. Rev. 27 (2), 121–138.
Soltow, L., 1981c. The distribution of property values in England and Wales in 1798. Econ. Hist. Rev. 34 (2), 60–70.
Soltow, L., 1985. The Swedish census of wealth at the beginning of the 19th century. Scand. Econ. Hist. Rev. 33 (1), 1–24.
Soltow, L., 1989. Distribution of Wealth and Income in the United States in 1798. University of Pittsburgh Press, Pittsburgh, PA.
Soltow, L., 1998. Inequality of income and wealth at the beginning of the 19th century. In: Soltow, L., van Zanden, J.L. (Eds.), Income & Wealth Inequality in the Netherlands 16th–20th Century. Het Spinhuis, Amsterdam.
Soltow, L., van Zanden, J.L., 1998. Income & Wealth Inequality in the Netherlands 16th–20th Century. Het Spinhuis, Amsterdam.
SOU, 1942. Förutsättningarna för och verkningarna av en engångsskatt å förmögenhet i Sverige. SOU 1942:52, Finansdepartementet, Stockholm.
SOU, 1957. Arvsbeskattning, betänkande av Arvsskattesakkunniga. SOU 1957:48, Finansdepartementet, Stockholm.
SOU, 1969. Kapitalbeskattningen, betänkande av Kapitalskatteberedningen. SOU 1969:54, Norstedt, Stockholm.
SOU, 2004. Egendomsskatter: reform av arvs- och gåvoskatter. Slutbetänkande av Egendomsskattekommittén. SOU 2004:66, Fritzes, Stockholm.
Spånt, R., 1979. Den svenska förmögenhetsfördelningens utveckling. SOU 1979:9, Fritzes, Stockholm.
Spånt, R., 1982. The distribution of household wealth in some developed countries. A comparative study of Sweden, Denmark, France, Germany, United Kingdom and United States. In: Kessler, D., Masson, A., Strauss-Kahn, D. (Eds.), Accumulation et Répartition de Patrimoines. CNRS and Economica, Paris.
Stamp, J.C., 1914. A new illustration of Pareto's law. J. Roy. Stat. Soc. 77, 200–204.
Stamp, J.C., 1916. British Incomes and Property: The Application of Official Statistics to Economic Problems. P.S. King, London.

Statistics Denmark, 1995. Indkomster og Formuer 1995. Copenhagen.
Statistics Denmark, 1996. Indkomster og Formuer 1996. Copenhagen.
Statistics Finland, 1911. Bidrag till Finlands officiella statistik, IV. Förmögenhetsförhållanden. Kungliga senatens tryckeri, Helsingfors.
Statistics Finland, 2006. Tulo- ja varallisuustilasto 2004. Statistics Finland, Helsinki.
Statistics Finland, 2007. Kotitalouksien varallisuus 1988–2004. Statistics Finland, Helsinki.
Statistics Netherlands, 2010. Procesbeschrijving statistiek Vermogens huishoudens, Vermogensstatistiek. Centraal Bureau voor de Statistiek, Den Haag/Heerlen.
Statistics Norway, 1915a. Indtaegs og formuesforhold efter skatteligningen 1911 i forbindelse med Folketællingen 1910, Kristiania.
Statistics Norway, 1915b. Indtaegt, formue og fordelingen av den kommunale skat ifølge skatteligningen for 1913–1914, Kristiania.
Statistics Norway, 1934. Folketellingen i Norge 1 desember 1930. 7 hefte. Inntekt og formue efter skatteligningen 1930–31, Oslo.
Statistics Norway (various years), Statistisk Årbok, Oslo.
Statistics Sweden, 1927. Folkräkningen den 31 december 1920. V. Yrkesräkningen, 2: yrke, inkomst och förmögenhet kombinerade inbördes samt med kön, civilstånd och ålder. P.A. Norstedt & Söner, Stockholm.
Statistics Sweden, 1937. Folkräkningen den 31 december 1930. VII. Folkmängden efter yrke, inkomst och förmögenhet: 2 avd. P.A. Norstedt & Söner, Stockholm.
Statistics Sweden, 1938. Folkräkningen den 31 december 1930. VIII. Folkmängden efter yrke, inkomst och förmögenhet: 3 avd. P.A. Norstedt & Söner, Stockholm.
Statistics Sweden, 1940. Särskilda folkräkningen 1935/36. VIII. Partiella folkräkningen i mars 1936: Yrke. Yrkesväxling. Skol- och yrkesutbildning. Inkomst och förmögenhet. P.A. Norstedt & Söner, Stockholm, Stockholm.
Statistics Sweden, 1951. Folkräkningen den 31 december 1945. VIII:1. Partiella undersökningar (bottensamplingen). Behandlar delar av statistiken över inkomst. Sveriges Officiella Statistik: Folkmängden och dess förändringar. K.L. Beckmans Boktryckeri, Stockholm.
Statistics Sweden, 1956. Folkräkningen den 31 december 1950. VII. Urvalsundersökningar. Statistiken över inkomst. Sveriges Officiella Statistik: Folkmängden och dess förändringar. Statistiska Centalbyrån, Stockholm.
Steinmo, S., 1993. Taxation and Democracy. Yale University Press, New Haven.
Stiglitz, J.E., 1969. Distribution of income and wealth among individuals. Econometrica 37 (3), 382–397.
Terviö, M., 2008. The difference that CEOs make: an assignment model approach. Am. Econ. Rev. 98 (3), 642–668.
Tinbergen, J., 1974. Substitution of graduate by other labor. Kyklos 27, 217–226.
Tinbergen, J., 1975. Income Distribution: Analysis and Policies. North-Holland, Amsterdam.
Tucker, R.S., 1938. The distribution of income among income taxpayers in the United States, 1863–1935. Q. J. Econ. 52 (4), 547–587.
Tuomala, M., Vilmunen, J., 1988. On the trends over time in the degree of concentration of wealth in Finland. Finnish Econ. Papers 1, 184–190.
U.S. Census Bureau, 2011. Income, Poverty, and Health Insurance Coverage in the United States: 2010. U.S. Census Bureau, Washington, DC.
van Zanden, J.L., 1998a. The egalitarian revolution of the 20th century. In: Soltow, L., van Zanden, J.L. (Eds.), Income & Wealth Inequality in the Netherlands 16th–20th Century. Het Spinhuis, Amsterdam.
van Zanden, J.L., 1998b. In conclusion. In: Soltow, L., van Zanden, J.L. (Eds.), Income & Wealth Inequality in the Netherlands 16th–20th Century. Het Spinhuis, Amsterdam.
van Zanden, J.L., 2009. The Long Road to the Industrial Revolution: The European Economy in a Global Perspective, 1000–1800. Brill, Leiden.
Veall, M., 2010. Top income shares in Canada: updates and extensions. Working Paper, McMaster University.
Venables, A.J., 2008. Rethinking economic growth in a globalizing world: an economic geography lens. Working Paper No. 18, The International Bank for Reconstruction and Development/World Bank, on behalf of the Commission on Growth and Development.

Vermaas, A., Verstegen, S.W., van Zanden, J.L., 1998. Income inequality in the 19th century: evidence for a Kuznets curve? In: Soltow, L., van Zanden, J.L. (Eds.), Income & Wealth Inequality in the Netherlands 16th–20th Century. Het Spinhuis, Amsterdam.
Waldenström, D., 2014. Wealth-income ratios in a small, late-industrializing, welfare-state economy: Sweden, 1810–2010. Working Paper, Uppsala University.
Watkins, G., 1907. The Growth of Large Fortunes: A Study of Economic Causes Affecting the Acquisition and Distribution of Property. Macmillan, New York.
Williamson, J., 1980. Earnings inequality in nineteenth-century Britain. J. Econ. Hist. 45 (3), 457–476.
Williamson, J.G., 2006. Globalization and the Poor Periphery Before 1950: The 2004 Ohlin Lectures. MIT Press.
Williamson, J.G., Lindert, P., 1980a. American Inequality: A Macroeconomic History. Academic Press, New York.
Williamson, J.G., Lindert, P., 1980b. Long-term trends in American wealth inequality. In: Smith, J.D. (Ed.), Modeling the Distribution and Intergenerational Transmission of Wealth. University of Chicago Press, Chicago.
Wilterdink, N., 1984. Vermogensverhoudingen in Nederland. Ontwikkelingen sinds de negentiende eeuw. De Arbeiderspers, Amsterdam.
Wolff, E.N., 1987. Estimates of household wealth inequality in the U.S. 1962–1983. Rev. Income Wealth 33 (3), 231–256.
Wolff, E.N., 1996. International comparisons of wealth inequality. Rev. Income Wealth 42 (4), 433–451.
Wolff, E.N., 2007. The retirement wealth of the baby boom generation. J. Monetary Econ. 54 (1), 1–40.
Wolff, E.N., 2012. The Asset Price Meltdown and the Wealth of the Middle Class. Unpublished manuscript, New York University.
Wolff, E.N., Marley, M., 1989. Long-term trends in U.S. wealth inequality: methodological issues and results. In: Lipsey, R.E., Tice, H.S. (Eds.), The Measurement of Saving. University of Chicago Press, Chicago, IL.
Wolff, E.N., Zacharias, A., 2009. Household Wealth and the Measurement of Economic Well-Being in the United States. J. Econ. Inequal. 7 (2), 83–115.
Wood, A., 1994. North-South Trade, Employment and Inequality: Changing Fortunes in a Skill-Driven World. Clarendon Press, Oxford.
World Bank, 2002. Globalization, Growth, and Poverty. World Bank and Oxford University Press, Washington D.C.
Zeuthen, F., 1928. Den Økonomiske Fordeling. Nyt Nordisk Forlag Arnold Busck, Copenhagen.
Zucman, G., 2013. The missing wealth of nations, are Europe and the U.S. net debtors or net creditors? Q. J. Econ. 128 (3), 1321–1364.

CHAPTER 8

Post-1970 Trends in Within-Country Inequality and Poverty: Rich and Middle-Income Countries

Salvatore Morelli[*,†], **Timothy Smeeding**[‡], **Jeffrey Thompson**[§]
[*]CSEF, University of Naples, Federico II, Naples, Italy
[†]INET Oxford, University of Oxford, Oxford, UK
[‡]University of Wisconsin, Madison, WI, USA
[§]Federal Reserve Board of Governors, Washington, DC, USA

Contents

Handbook of Income Distribution, Volume 2A
ISSN 1574-0056, http://dx.doi.org/10.1016/B978-0-444-59428-0.00009-6

Abstract

Like the other chapters in this volume of the *Handbook of Income Distribution* (and its predecessor), the aim of this chapter is to provide a comprehensive review of a particular area of research. We examine the literature on post-1970 trends in poverty and income inequality, up to 2010 or 2011 in most countries. We provide measures of the levels and trends in each of these areas, as well as an integrated discussion of empirical choices made in the measurement of poverty, overall income inequality, and inequality among those with top incomes.

JEL Classification Codes

D30, D31, D63, H23, I30, I32, N30, O15

Keywords

Trends in income inequality, Poverty, Gini, Top income shares, Inequality

8.1. INTRODUCTION

The chapter on "Post-1970 Trends in Within-Country Inequality and Poverty" was rather sketchy 13 years ago when the first volume of the *Handbook of Income Distribution* was published (Gottschalk and Smeeding, 2000). A separate chapter was devoted to poverty measurement (Jäntti and Danziger, 2000). The first Canberra Report (2001) on international standards for income distribution, much less the second report (2011), had not yet been published, and the Luxembourg Income Study (LIS) database had less than 15 years of comparable data on poverty and inequality for fewer than 20 rich nations, for which only trend data from 1980–1995 were available. The Organization for Economic Cooperation and Development's (OECD) method for collecting comparable income distribution data was in its infancy, and the top incomes database was even younger still; Piketty (2001) was just publishing his paper on the long-term distribution of top incomes in France.

The data world has come a long way in 14 years; now available are multiple sources of comparable (harmonized) household income data (overall and top incomes), wealth data, and poverty data. Even heeding warnings to take caution with harmonization of secondary data (Atkinson and Brandolini, 2001), the world now has a substantial number of more comparable data series, both across nations and over time. Still, there are limits to what can be accomplished in terms of comparisons. Here we rely on high-quality comparable level and trend data for income and income poverty from the 1970s on. While our measures dwell heavily on OECD countries, using both the LIS and the most recently available OECD data, particularly for the richest of these nations, improved availability has allowed us to add some data for "developing" nations and middle-income countries (MICs).[1]

Two other chapters in this volume cover longer-term trends in inequality and developing country inequality (Chapters 7 and 9, respectively). Chapter 23 focuses on the effects of policy on poverty. We overlap to a small degree with these chapters. As argued below, our income and poverty measures are based on commonly defined and measured disposable income as well as pretax income for tax records. Consumption data are not yet comparable enough to use in cross-national analysis; wealth data comparability has begun but not yet flowered. There are only scattered cross-national studies of wealth or asset poverty. While the European Union (EU) and some OECD studies consider indices of material deprivation, such measures are not standardized enough in the rest of the world to be examined here, as elaborated in the next section, where we select yardsticks.

Next we turn to income poverty measures, mainly those based on data from the LIS and the OECD, where we examine levels, trends, and both anchored and relative poverty. We then turn to the topic of overall income inequality levels and trends, based on these same sources, before turning to top-end inequality measures. In so doing we attempt to bridge the divide between household income distribution data, from surveys and registers, and top income data based on taxable income and income tax units, highlighting the extent of complementarity and substitutability between the two. The final section summarizes our review and the conclusions we draw.

8.2. CHOOSING A YARDSTICK AND ITS COMPONENTS

Multiple perspectives can be used to evaluate the distribution of living standards in a society. These focal points, as labeled by Sen (1992, p. 20), include monetary indicators such as expenditure, income, and wealth, as well as nonmonetary indicators such as multidimensional measures of material standard of living, happiness and life satisfaction, functioning, and capabilities. Here disposable income is taken as the focal variable for overall inequality and poverty trends, and taxable income records for tax filing units which permit long-term and accurate investigations of the incomes of the top strata of society.

The distribution of income among persons, or households, has attracted the attention of social scientists at least since Gregory King's 1688 social tables, "which offer unique quantitative views of social structure and income distribution during a statistical Dark Age" (Lindert and Williamson, 1982). Pareto's analysis of the revenue curve in 1897

[1] We include Brazil, Bulgaria, China, Chile, India, Hungary, Romania, South Africa, Turkey, and additional nations in some of the analysis, based on LIS data (http://www.lisdatacenter.org/) or OECD data (www.oecd.org/social/income-distribution-database.htm).

is a more recent formalization of this interest. Income is still the most common indicator of economic resources in rich countries. While consumption expenditure is often used in developing countries, the Hicks–Hansen identity for income (or potential consumption), which is equal to actual consumption plus the change in net worth[2] over a given period, ideally ties income and consumption neatly together. But no one data set contains fully comparable measures of all three ingredients in any nation, mainly because change in net worth is difficult to measure (see also Brandolini and Smeeding, 2009; Fisher et al., 2012).

8.2.1 Consumption or Income?

The nearest alternative to disposable income is consumption or consumption expenditure, a variable that is often preferred in less developed countries because it is more easily measured in such localities. Consumption can be smoothed over time and therefore is less volatile and less reliant on seasonal variation than is income, especially in agricultural societies (Deaton and Grosh, 2000). Apart from this practical reason, many economists view consumption as a better proxy of well-being than income (Fisher et al., 2012). One argument is that well-being (utility) is a function of the goods and services actually consumed, not those merely owned (Slesnick, 1994). However, focusing on the means available to purchase commodities (income) rather than the commodities actually purchased (expenditure) makes the assessment of well-being independent of the purchase choice. Sen (1992) offers the example "... of the person with means who fasts out of choice, as opposed to another who has to starve because of lack of means" (pp. 111–112), whereas Hagenaars and colleagues (1994, p. 8) argue that using income helps us avoid the trap of confusing voluntarily low levels of consumption with material deprivation.

A second argument in favor of consumption is that it is more closely related to permanent income or lifetime resources than current income. As described by Friedman (1957, p. 209), the distributions of current income "... reflect the influence of differences among individual units both in ... the permanent component of income and ... the transitory component. Yet these two types of differences do not have the same significance; the one is an indication of deep-seated long-run inequality, the other, of dynamic variation and mobility." If one is interested in "deep-seated long-run inequality," permanent income and, hence, consumption are what matter. However, the simple proportionality between consumption and permanent income in the baseline intertemporal consumer's optimization problem does not hold if some of its basic hypotheses are relaxed and simple forms of personal heterogeneity are introduced (effects of accumulated or inherited wealth, the degree of intergenerational altruism, the variability of uncertain labor incomes, and capacity to borrow, to name just a few). Therefore, current

[2] Change in net worth can be positive (net income saved) or negative (net debt incurred).

consumption may not be a very good, and not even the best available, proxy of permanent income. Moreover, it is far from obvious that "deep-seated long-run inequality" should be our major concern. The concept has some natural appeal: an undergraduate may have a current income below that of a manual worker of the same age, but she is likely to be better off within a few years and for most of her lifetime. But "the promise of resources in the future may do little to pay the bills today" (Deaton and Grosh, 2000, p. 93). In the real world, capital markets are imperfect, and units face borrowing constraints that render the actual standard of living dependent on currently available resources. Conversely, "... the fact that an old person had a high income thirty years ago does not make up for his having a pension that is below his needs today" (Atkinson, 1983, p. 44).

Finally, there is the problem of measuring "true" consumption in rich societies. Consumption expenditure data are collected mainly to provide weights and prices for measuring the Consumer Price Index, not for measuring consumption. Few surveys actually try to measure actual consumption because purchases of durables such as major appliances, automobiles, and especially housing must all be spread out over the useful life of the good, which is bought in one period but consumed in another. Indeed, measures of consumption may differ greatly from consumer expenditures for such persons as older individuals living in an owned but mortgage-free house (Fisher et al., 2012; Johnson et al., 2005; Meyer and Sullivan, 2012a,b).

In brief, there is *a priori* no cogent or practical reason to prefer consumption to income or permanent income to current income. Indeed Haig (1921) and Simons (1938) recognized that income represents the possibility to consume and therefore established their famous identity that income equals consumption plus or minus changes in net worth. Most often, the choice is driven by available information, and there is a clear preference among rich nations to rely on income and not consumption. MICs also are increasingly likely to have living standards better measured by incomes, especially in their rapidly growing urban areas. Indeed, if the value of informal labor is captured (including production for personal consumption) then income and consumption differ only by changes in net worth, which may be small in the less modern regions of MICs. Our income data on MICs, presented below, are based on such a definition of income. Current income, therefore, seems to be a satisfactory measure of people's (material) living standard.

After settling on income as the focal variable, however, a number of important conceptual issues and data concerns remain. In addition to the issues of data availability over time and comparability across countries, the analysis of distributional measures requires decisions and assumptions regarding the income concept, the income-sharing unit, the accounting period, and statistics for measuring poverty, material hardship, or the distribution of income (Johnson and Smeeding, 2013; Smeeding and Weinberg, 2001).

8.2.2 The Definition of Income and Other Essentials

The most basic income concepts collected by national statistical agencies and used by researchers are market (factor) or pretax and transfer income and disposable income. On the basis of the recommendations of the reports of the Expert Group on Household Income Statistics—The Canberra Group (2001, 2011), market income should include all types of earnings gross of employees' social insurance contributions; net self-employment income[3]; all types of capital income, including interest, rent, or dividends received (but not accrued); and subtracting interest paid and adding private pensions.

Disposable income takes market income and subtracts direct taxes (including an employee's contributions to social insurance) but ignores other "indirect" taxes (property, wealth, and value-added taxes); then it adds back in regular interhousehold cash transfers received net of those made, as well as all forms of cash and near-cash public income transfers including social insurance benefits (for social retirement, disability, and unemployment); universal social assistance benefits; and targeted income transfer programs such as social maintenance. Near-cash benefits in the form of housing allowances or food stamps are included, as are negative taxes (for instance, in-work benefits now popular in many rich nations).[4]

However broad these definitions might be, they exclude imputed rents, capital gains and losses, and other unrealized types of capital income, home production, and in-kind transfer benefits such as education and health insurance. Because these items may account for an important share of the economic resources at the household's disposal, their inclusion in the income definition may affect measured inequality. Indeed, research on the United States suggests that uncounted realized and unrealized income from capital increases measured incomes by over 40% at the mean and more than 20% at the median (Smeeding and Thompson, 2011).

Imputed rent for owner-occupied dwellings tends to benefit a wide range of low- to high-income units, especially the elderly, but their overall effect may vary across countries, depending on the level of housing prices and the diffusion of home ownership (Frick and Grabka, 2003). Unrealized appreciation and untaxed income from capital, as well as capital gains, mainly benefit higher-income units. Indirect taxes have a relatively

[3] All surveys net out the costs of producing self-employment income, but none also deduct the costs of earning wage income, such as child care. More generally, traditional household income data cannot account for the cost of foregone home production, especially when parents switch their work from household income to market earnings, as has taken place widely in rich countries since the 1970s. Because of the change in modes of production, and the failure to account for the direct and opportunity costs of earning incomes, the rise in secondary earners among partners with children probably overstates their net income gain when both partners work.

[4] In practice, many surveys also exclude various elements of market incomes, such as interest paid or private transfers made to other households, and therefore these are often ignored.

larger impact on the budget of lower-income units (Newman and O'Brien, 2011), but the opposite happens with the imputation of in-kind public benefits for health care, housing, and education valued at their cost of provision. Because the value of these benefits is spread more or less evenly among beneficiaries ("potential" beneficiaries in the case of health insurance), the typical approach is to augment income by a fixed amount, which accounts for a larger fraction of income at lower-income levels (Burkhauser et al., 2012b). In general, elder households and households with children are net gainers from the imputation through health insurance and education benefits, respectively, whereas middle-aged childless units are net losers (Garfinkel et al., 2006, 2010). These results are very sensitive to the imputation assumptions: both valuing benefits according to willingness to pay and accounting for the quality of services provided would reduce benefits to the poor (Smeeding, 1982).

As stressed in the first Canberra Group Report (2001, pp. 62–67), the undercoverage of property and self-employment income, own account production, imputed rent for owner-occupied dwellings, in-kind social transfers, capital gains, and other unrealized income from wealth are major issues to be addressed in expanding internationally comparable income measures. But the analysis of these augmented notions of income is also scarce at the national level. Despite these omissions and shortcomings, market income and disposable household income (DHI) remain the standard concepts measured and published by national statistical agencies and research institutions.

8.2.2.1 Reference Period, Income Units, and Resource Sharing

Income is a flow of resources received by people over a given period. To have a coherent concept of income that can be used to compare distributions across countries and analyze trends over time requires common units to describe the period over which income is received and the groups of people who are sharing the income.

The statistics and trends analyzed in this section are all based on annual data, in part because of convention and data availability. The choice of the reference period does, however, have implications for the degree of inequality in the distribution measured at any given point in time. In the presence of fluctuations in income, where some households experience positive or negative shocks or lumpy income streams, the distribution of income will seem more unequal the shorter the reference period (Atkinson, 1983; Atkinson et al., 1995). At intra-annual frequencies, income may fluctuate because of seasonal factors (e.g., in agriculture), movement of workers into or out of jobs, or the timing of payments (e.g., interest on financial assets or liabilities, dividends on stocks). Aggregating over the year implies averaging out these differences, although the overall impact on measured inequality may be small (Böheim and Jenkins, 2006). By the same token, lengthening the reference period beyond the year reduces measured inequality by smoothing the variability due to the business cycle or the life cycle (e.g., Björklund, 1993; Björklund and Palme, 2002). Longer periods of time may come closer to approximating the "lifetime income" concept preferred by some economists, but in practice these data are quite rare. Using Swedish data, Björklund (1993) found that the dispersion of four decades' worth of cumulative income data for individuals was up to 40% lower than dispersion measured from a standard cross section.

Income is typically shared across family or household units. Analysis of the distribution of income across countries and over time requires both adjustments for the economies of scale associated with income sharing and the use of comparable income units. The typical income-receiving unit is the household, but some data sources report income for individuals, families, or tax-paying units, which potentially include individuals, families, and subfamily units. The broader the definition of household, the more measured inequality tends to decrease, because the dispersion of individual incomes is abated by their aggregation and supposedly egalitarian distribution among all members of the unit (Redmond, 1998). The poverty trends discussed below in Section 8.3 and the distributional measures for the entire population discussed in Section 8.4.1 are based on household income surveys and use the household as the income unit. The trends in high-income shares discussed in Section 8.4.2 are typically based on tax-paying units; they are commonly based on national income tax statistics, and multiple tax units may be included in one household.

It is widely accepted that there are greater costs associated with larger households and economies of scale in consumption that are generated by cohabitation. A family with two children faces greater costs than a family with one child, with greater expenses for food, clothing, education, transportation, and housing. As a result, the same level of after-tax income implies a lower material standard of living for the larger family. With economies of scale in a household, though, providing for the second child will not be as costly as providing for the first. Similarly, a couple living together will spend more on housing, utilities, food, and transportation than a single person, but the couple does not need to spend twice as much to obtain the same standard of living, all else being the same.

To account for costs associated with household size and the related economies of scale, researchers have developed different "equivalence scales" to create comparable incomes of different household sizes and compositions. The most commonly used equivalence scale, taken from Buhmann et al. (1988), further described by Atkinson et al. (1995), and recommended by the Canberra Group, divides household income by the square root of the household size. Using the square root scale, costs increase with the household size, but at a declining rate. The square root scale, though, does not explicitly acknowledge differences in the cost of living between adults and children. The LIS project uses the square root scale, and the OECD has used it in its publications since 1995. The EU uses an alternative scale when calculating distributional statistics with the Statistics on Income and Living Conditions (SILC) data (Atkinson et al., 2010a,b). The scale used by the EU divides household income by the weighted number of household members; different weights are applied to adults and children. The household head is given a weight of 1, each additional adult household member a weight of 0.5, and each child a weight of 0.3.[5] The U.S. Census Bureau adopted a three-parameter equivalence scale that further differentiates between children in different household types. The census scale, discussed by Short (2001), reflects the idea that children in single-parent families represent a greater increase in costs than do children in two-parent families.

The choice of the equivalence scale affects inequality comparisons. It also affects poverty comparisons, especially between those who typically live in small units (elderly) or larger units (families with children or multigenerational units) (Buhmann et al., 1988; Coulter et al., 1992).

Finally, the welfare weighting of the single observations may vary. Each observation may receive a weight of 1 (household weight) or may be weighted according to its size (person weight) or its size and composition (equivalent adult weight), again bringing differences in poverty and inequality outcomes (Danziger and Taussig, 1979; Ebert, 1997).

8.2.3 Data Source Comparability: Surveys, Tax Records, and the Rich

The last cause of limited comparability may be attributable to differences in the source of data. Income data are available both from national household surveys and from administrative archives. Of the latter, the most important are income tax records, which have historically provided long runs of continuous data and have been exploited in the literature on top incomes (Atkinson and Piketty, 2007). Income tax records suffer from potentially serious problems, including the incomplete coverage of those with incomes below the tax threshold, inability to adjust for household size, and the tendency to underreport certain types of income. These and other methodological issues related to tax records and calculation of top income shares are discussed in greater detail in Section 8.4.2.

Household surveys are also subject to problems, including sampling errors, which depend on the size and structure of the sample, and nonsampling errors caused by nonresponse and underreporting (see Chapter 2 of Atkinson et al., 1995). For these reasons, the upper tail of the income distribution tends to be unsatisfactorily covered in sample surveys, unless the rich are oversampled and reporting errors are minimized. The survey-based evidence discussed later in this chapter may be seen as being about the incomes of the bottom 95–99% of the population, and it is thus complementary but not always fully comparable to the results of high incomes based on tax records reported

[5] The equivalence scale used in the Eurostat figures is sometimes referred to as the "modified" OECD equivalence scale because it supersedes another scale previously used by the OECD (see http://www.oecd.org/els/soc/OECD-Note-EquivalenceScales.pdf).

in the final section of this chapter.[6] The specific statistics used in the calculation of poverty and income inequality using household survey data are discussed below in Sections 8.3 and 8.4.1.

All these factors need to be kept in mind in the analysis of the national trends in income inequality or in cross-national comparisons. While the data include a great deal of "noise" or possibly unknown errors, the important assumption is that the signal derived from the analysis exceeds the noise for most careful analyses, which also include sensitivity tests of assumptions (Atkinson et al., 1995; Gottschalk and Smeeding, 2000). In examining trends, we are aided by the fact that errors may be more consistent across multiple rounds of the same survey, and therefore trends may be more cross-nationally reliable and comparable than levels of inequality (Gottschalk and Smeeding, 1997). But even then, almost all surveys undergo often substantial changes over multiple decades, producing artificial changes in results due to changes in sampling, survey mode, or other changes in procedures.

Finally, full comparability is an impossible goal. Surveys within countries as well as across countries are subject to changes in methods and are characterized by differences in sampling and nonsampling errors. Comparability is vastly increased when the researcher can access the individual observations on household incomes available in a national archive or in international databases, where the original databases are harmonized, such as the LIS and the EU-SILC. Here, both levels and trends are more comparable than using other methods. Ex ante instructions to compute a series of harmonized data also are available from the OECD (2008, 2011, 2013).

Since 1983, the LIS Cross-National Data Center has been creating "harmonized" income data sets for a growing number of countries. LIS works with the existing income surveys of different countries and converts them to a format with consistent definitions and concepts that make cross-national comparisons possible. By way of nondisclosure agreements and secure remote access servers, LIS also makes possible for research access to income surveys from a number of countries that traditionally do not share their underlying data. By 2012, LIS included eight different waves of harmonized data covering roughly equivalent points in time between 1967 and 2010 across countries. The initial LIS wave included 7 countries, but the number has grown steadily, reaching nearly 40 countries in the most recent waves.

The EU's statistical agency, Eurostat, provides comparable income survey statistics for the EU member countries. Eurostat initially used a common survey instrument across the European counties but has since switched to an "ex-ante harmonized" framework (Atkinson et al., 2010a,b). The European Community Household Panel Survey covered 15 different countries from 1994 to 2001 and was replaced by the SILC. SILC works

[6] Data may be bottom and top coded, either in the course of data collection, as in the U.S. Current Population Survey (Ryscavage, 1995), or as a decision of the researcher to reduce the noise that is typically concentrated in the tails of the distribution (Burkhauser et al., 2009; Cowell and Victoria-Feser, 1996).

through the statistical agencies of the different EU member countries and achieves cross-national comparability through adoption of common definitions and concepts key to cross-national comparability of income and other policy-relevant matters in the EU.[7] In 1995 there were 13 countries initially represented, but the number of countries expanded to 22 by 2000 and 30 by 2005. The income distribution measures produced by Eurostat now cover 32 different countries. In contrast to LIS, the Eurostat distribution statistics are produced annually; covering the years between 1995 and 2011 there are 380 year-country observations for the different distributional measures.

The OECD also regularly releases income distribution and poverty measures for its member countries. These releases have been highlighted in major publications, including *Growing Unequal?* (2008), *Divided We Stand* (2011), and *Crisis Squeezes Income and Puts Pressure on Inequality and Poverty* (2013), and are also available in the organization's Household Income Distribution and Poverty online databases (www.oecd.org/social/inequality.htm). The OECD figures are based on the national statistical agency household surveys and are created by a network of country specialists using common measures.[8] Since the figures are calculated from country-specific surveys in different years, the data are not always based on the same years. In several of these publications, the OECD data compare fairly well with the LIS data observed in the same year, but then the OECD methods add more up-to-date data than those available from the LIS. In the mid-1970s, 8 countries were represented in the distributional statistics, but by the late 2000s the number of countries had grown to 34. Because these data tend to be more immediate and can be updated with less ex-post harmonization than, say,

[7] Even within this common framework, there are important differences in the approaches the different EU countries use in their SILC income surveys. See Iacovou et al. (2012) for a detailed discussion of some of the weaknesses of the EU-SILC data and potential problems for using these data for cross-national analysis. Several EU countries (Finland, the Netherlands, Norway, Slovenia, and Sweden) use administrative records supplemented with interviews of representative household members. Most of the countries use rotational panel household surveys, but there is considerable variation in the number of rotation groups and length of time in the panel. In most of the countries the length of the panel is 4 years, with one rotation group dropped every year, but Norway and France have 8- and 9-year panels, respectively, and Luxembourg uses a traditional panel. Spain and Ireland use substitutes for nonresponders in their household survey. Before 2008 Germany used a combined quota and random sample for its survey.

[8] At least some of the surveys used by both OECD and LIS are household surveys combined with samples drawn from administrative registers in the Nordic nations, Austria, France, and Denmark. These data suffer from less item nonresponse and reporting error than most household surveys, and they include a full random sample of all households, including the top 1%, thus also improving nonresponse and data quality. Because so much of the income gain in the past two decades has been at the top end of the distribution, these national register data might have become systematically different from the normal survey data in these databases. See OECD (2012) for a list of their surveys based on registers compared with other types of sample surveys. See also Chapter 2 of Atkinson et al. (1995) for a more general discussion.

LIS, we use OECD (2013) poverty data to capture the effects of the Great Recession (GR) on poverty below.

8.3. POVERTY MEASUREMENT AND TRENDS

In this section we examine the complexities of poverty measurement from its origins to current practice. We rely mainly on the LIS and OECD data to examine levels and trends in overall poverty, but we also refer to the literature on child and elder poverty. In our empirical examinations we look at both rich countries and MICs, comparisons of trends in relative poverty over different time periods, comparisons of relative and anchored poverty across the GR, and finally the correlation between levels of relative poverty and inequality as an introduction to Section 8.4 on overall income inequality.

8.3.1 Origins and Development of Poverty Measurement

The fundamental concept of poverty concerns itself with having too few resources or capabilities to participate fully in a society. As Blank (2008) reminds us, "poverty is an inherently vague concept and developing a poverty measure involves a number of relatively arbitrary assumptions." Ultimately, social scientists first need to establish the breadth and depth of this social phenomenon called "poverty" before they can meaningfully analyze it and explore its ultimate causes and remedies. Thus, we turn to measures and comparisons of poverty used by economists and other social scientists within and across nations.

Our discussion is framed by Figure 8.1, which reviews most of the possibilities of poverty concepts and measures. Here we are mostly interested in the concept of objective poverty measures, according to some standard definitions of means versus resources.[9]

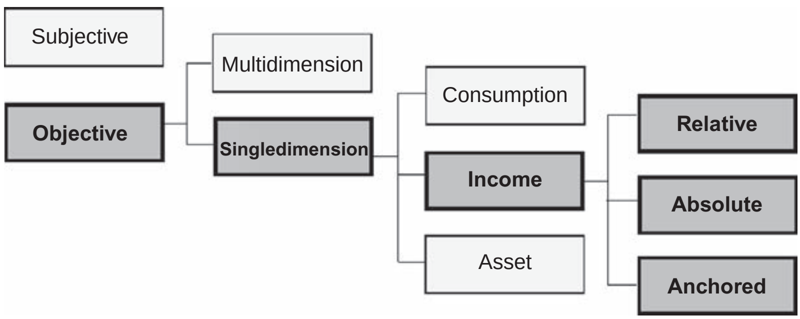


Figure 8.1 Conceptualizing poverty and its measurement. *Source: Adapted from Dhongde (2013).*

[9] See also Chapter 22, "The Idea of Antipoverty Policy" by Ravallion (2014) on the origins of poverty in economic thought and the role of antipoverty policy and social protection in meeting basic human needs.

In this chapter we are mostly interested in objective poverty measurement using a single dimension of "resources," income, and several notions of "needs" standards: those that are relative, absolute, and closely related anchored poverty lines. We chose income poverty because of its domination in modern (post-1960s) poverty studies and because of its linkages to the income inequality literature that follows. Other measures and concepts of poverty also are mentioned but not empirically investigated.

Income or living standards poverty measurement began in the Anglo-Saxon countries and dates back at least to Rowntree (1901), who was the first to use the concept of a poverty line in his empirical work on York, England. Thanks to his enterprise, and that of Booth (1903), who invented the idea of a poverty line for London, we have a meaningful social indicator of basic needs (see, for instance, Piachaud, 1987; Ravallion, 2014; Ringen, 1985; Townsend, 1979, 1993). We also note that official poverty measurement began as an Anglo-American social indicator. Since then, "official" measures of poverty (or measures of "low income") now exist in over 100 countries and for Europe as a whole (Eurostat, 2005). The United States (DeNavas-Walt et al., 2012) and the United Kingdom (Department for Work and Pensions, 2012) have long-standing "official" poverty series. Statistics Canada publishes on an irregular basis the number of households with incomes below a "low-income cutoff", as does the Australian government with those below the "Henderson line." In Northern Europe and Scandinavia the debate centers instead on the level of income at which minimum benefits for social programs should be set. In other words, their concept of insufficient "low income" is directly fed into programmatic responses to social needs (Björklund and Freeman, 1997; Marx and Nelson, 2013; Ravallion, 2014).[10]

While poverty measurement is an exercise that is particularly popular in English-speaking countries, most rich nations share the Anglo-Saxon concern over distributional outcomes and the well-being of the low-income population. There is no international consensus on guidelines for measuring poverty, but international bodies such as the United Nations Children's Fund (UNICEF, 2000), the United Nations Development Programme (UNDP, 1999), the Organization for Economic Cooperation and Development (OECD, 2008, 2013), and the European Statistical Office (Eurostat, 1998, 2005) have published several cross-national studies of the incidence of poverty in rich countries. The large majority of these studies have been based on the LIS database, which can be accessed at www.lisdatacenter.org. Some examples of these studies include Förster (1993), Jäntti and Danziger (2000), Smeeding et al. (2000), Kenworthy (1998),

[10] In addition to these objective poverty measures, several economists have used subjective measures of poverty and well-being, including income sufficiency (Groedhart et al., 1977; Hagenaars and van Praag, 1985; Ravallion, 2012; van Praag, 1968).While we do not cover these types of measures here, they are much akin to the chapter on subjective well-being and inequality later in this volume (Clark and D'Ambrosio, 2014).

Smeeding et al. (1990), and Smeeding (2006). More recently the European Union and the OECD have regularized measurement of poverty but using different standards and data sources. Today one can find poverty measures in over 100 countries, as well as some harmonized measures from the World Bank that use both secondary (published) data- and microdata-based measures of consumption and income to determine those living below some particular amount of income per person day, from $1.25 to $2.00 (Chen and Ravallion, 2012; Ravallion and Chen, 2011b).

8.3.2 Measuring Poverty

Most broadly, the measurement of poverty in rich nations involves the comparison of some index of household well-being with household needs. When command over economic resources falls short of needs, a household (or person or family) is classified as poor. Well-being refers to the material resources available to a household. Among most social scientists, the concern with these resources is generally not with material consumption *per se*, but rather with the capabilities such resources give to household members so they can participate fully in society (Brandolini and Smeeding, 2009; Sen, 1983, 1992). These capabilities are inputs to social activities, and participation in social activities gives rise to a particular level of well-being (Coleman and Rainwater, 1978; Rainwater, 1990). Methods for measuring a person's or household's capabilities differ according to the context in which one assesses them, either over time or across nations or among subpopulations within a nation, for example, rural versus urban China.

All advanced societies are highly stratified; hence, some individuals have more resources than others. The opportunities for social participation are affected by the resources that a household disposes, particularly in nations such as the United States, where there is heavy reliance on the market to provide such essential services as health care, postsecondary education, and child care. Monetary income is therefore a crucial resource. Of course, there are other important kinds of resources, such as social capital, wealth, noncash benefits, primary education, and access to basic health care, all of which add to human capabilities (Coleman, 1988). These resources may be available more or less equally to all people in some societies, regardless of their monetary incomes. There are many forces in rich societies that reduce well-being by limiting capabilities for full participation in society, including inadequacies in neighborhoods where people live, racial and ethnic discrimination, neighborhood violence, low-quality public schools and other social services, lack of good jobs, and job instability, all of which increase economic insecurity, reduce human capabilities, and increase poverty.

Because there is no single commonly accepted way to measure poverty among social scientists, there is a desire to go beyond the popularly used income poverty definition used below. So there exists a wide variety of additional poverty measures that substitute for or complement the preponderance of income-based measures used by quantitative sociologists and economists (see, e.g., Boltvinik, 2000; Haveman, 2009; Ruggles, 1990). In principle, poverty is a multidimensional concept and should reflect several aspects of personal well-being, as shown in Figure 8.1. Forms of deprivation other than economic hardship can certainly be relevant to poverty measurement and to antipoverty policymaking. A number of authors have suggested that separate measures of needs ought to be developed for different goods and services (Aaron, 1985). Housing and health care often are mentioned in this context, although the latter is particularly of interest in medically unequal nations such as the United States, whereas the former is of much greater interest in the United Kingdom (United Kingdom Department of Social Security, 1993).

The concept of multidimensional poverty is also flourishing. Official measures of social exclusion, material deprivation, and material hardship exist mainly in Europe, although they are beyond the empirical bounds of this chapter. Europe adopted the official Laeken set of social indicators in 1995, including the at-risk-of-poverty indicator, with an explicit objective of reducing poverty and social exclusion (Marlier et al., 2007). Indeed, indicators of material deprivation now form part of the Europe 2020 target of poverty reduction (Atkinson and Marlier, 2010, Chapter 6).

Both consumption poverty and asset poverty have been proposed as an alternative to income poverty in rich nations (Brandolini et al., 2010; Meyer and Sullivan, 2012a,b). In a few nations, asset and income poverty can be combined into a joint measure (Gornick et al., 2009), as can consumption and income poverty (Meyer and Sullivan, 2012a,b). But consumption and asset poverty measures are not yet ready for widespread use on a cross-national basis, despite their usefulness for some types of poverty measurement (e.g., many income-poor elderly consume more than their incomes because of dissaving and spending from assets).

In summary, we are interested primarily in comparative cross-national poverty measured in terms of income, not only because income-based poverty measures are more comparable across nations but also because income-based poverty allows us to connect our empirical work to overall inequality *per se* in the rich nations and MICs observed in this chapter. As mentioned above, income is generally a better measure of resources than consumption in rich countries. In the rapidly growing MICs, the differences in living standards between rural and urban populations cause the most angst over consumption versus income poverty. The richer the country, however, the more income becomes a better and more comparable measure. At the frontier of such comparisons, work by the LIS on "production for own consumption" and "informal labor" income help ease the comparisons across diverse areas within nations.

8.3.3 Measuring Absolute, Relative, and Anchored Poverty in Rich and Selected MICs

An absolute poverty standard is defined in terms of a level of purchasing power that is sufficient to buy a fixed bundle of basic necessities at a specific point in time. A relative standard, on the other hand, is defined relative to the typical income or consumption level in the wider society. The purchasing power of a relative poverty standard changes over time as society-wide income or consumption levels change, whereas an absolute poverty standard changes only with the prices of commodities it can buy. Most cross-national comparisons use the relative definition of poverty, especially because purchasing power parities to convert any absolute measure to country currency are subject to fluctuation and sometimes severe measurement error (Jäntti and Danziger, 2000).

In the broadest sense, all measures of poverty or economic need are relative because context is important to the definition of needs. The World Bank uses poverty measures of \$1.25–2 per person per day—or \$1095–2190 per year for a family of three—for the developing nations of Africa, Central Asia, or Latin America (Chen and Ravallion, 2012). In contrast, the 2011 United States "absolute" poverty threshold was about \$18,000 for a family of three—8–17 times the World Bank's poverty line; one-half of median income, the preferred relative poverty standard in the United States, is another 25% above this poverty line, or 10–21 times the poverty standard in poor countries. Moreover, as economic inequality has increased in most rich societies over the past 20 years, the study of relative deprivation and poverty has taken on new life (Gornick and Jäntti, 2013; Gottschalk and Smeeding, 2000; OECD, 2011, 2013).

Cross-national comparisons of poverty in rich countries therefore rely heavily on relative concepts of poverty, which are a reflection of the fact that a poverty standard or a minimum income standard ought to reflect the overall standard of living in society. One early source of this formalization (Abel-Smith and Townsend, 1965) came about in arguing that the officially defined minimum level of income in the United Kingdom, as represented by the National Assistance scale, should increase with the rising standard of living, and not just with consumer prices. It was Townsend's work in the early 1960s, culminating in his famous 1979 book, that really launched the relative poverty approach on a much wider scale.

As Townsend (1979, p. 31) wrote:

> *Individuals, families and groups in the population can be said to be in poverty when they lack the resources to obtain the type of diet, participate in the activities and have the living conditions and the amenities which are customary, or at least widely encouraged or approved in the societies to which they belong.*

The measurement of relative poverty has more recently been generally operationalized with a definition of the poverty line as a fraction of median income. Cross-national studies typically compare the percentage of people living with income below some fraction of the family-size-adjusted national median income.

Measurement of relative poverty in the United States also began in the 1960s and was pioneered by Fuchs (1967), who followed the thinking of Townsend and Abel-Smith and linked relative and absolute income poverty measurement. When Fuchs began his study, the absolute poverty measure in the United States begun by Lampman (1964) and then Orshansky (1965) was based on a poverty line of about $3000 for four persons. Fuchs pointed out that this was half of the median income at that time and that one could think differently about relative poverty compared with absolute poverty (Gilbert, 2008, p. 136).[11]

A relative poverty measure comparison is also consistent with a well-established theoretical perspective on poverty (Sen, 1983, 1992; Townsend, 1979). However, the fraction of income at which the poverty line ought to be set is open to debate. Most cross-national studies (LIS, OECD) focus on half of the median income, following Fuchs and others. But many feel that a 50% of the median standard is too low. It implies a poverty cutoff well below half the mean in unequal societies,[12] and it also affects the country rankings.[13] The European Statistical Office Working Group on Poverty Measurement has used 60% of the national median income as the common poverty threshold for European Community poverty studies in the new millennium (Eurostat, 2005, 2011).

A fully relative measure of poverty changes in lock step with median income, whereas an absolute measure changes only with prices. The income elasticity of the poverty line is therefore between 0 for the absolute measure and 1 for the fully relative measure. In some countries, such as the United States, the measure of poverty has become "semi-relative" as the poverty line advances only with the living standards of the bottom part of the distribution and not the whole distribution (Short, 2012). Ravallion and Chen (2011a) refer to "weakly relative measures," which have the feature that the poverty line will not rise proportionately to the median or mean, but will have income elasticity less than unity. These are also-called quasi-relative poverty standards in the new "Supplemental Poverty Measure" for the United States, which varies by considering expenses on basic needs for a low-income family and how they change over time (Short, 2012).

Understanding both absolute and relative poverty measures is worthwhile because they tell different things about living standards as well as deprivation. Increasingly, the idea of "anchored" poverty measures have become important and can be used to indicate both relative (or weakly relative) and absolute poverty trends within a given nation.

[11] Lampman's chapter, "The Problem of Poverty in America" (part of the 1964 Economic Report of the President), preceded President Johnson's declaration of the "War on Poverty" in his 1964 State of the Union Address. But while Lampman used $3000 of monetary income for his measure, it was not adjusted for family size. Orshansky (1965) produced a measure that had a similar poverty count and a similar poverty line for a four-person family but that differed by family size. In the late 1960s, Orshansky's measure became the official measure of poverty in the United States.

[12] Most relative poverty or deprivation measures rely on the median, not the mean income, especially in cross-national studies, because the latter may be affected by sampling and nonsampling error in different surveys. Moreover, the reference to the standard of living enjoyed by the middle or average family means the median family. See Smeeding et al. (1990).

[13] See LIS key figures for country poverty rates at 40, 50, and 60% of the median at http://www.lisdatacenter.org/data-access/key-figures/search/) and compare rankings there with the European Community rates in 2012 as taken from the Eurostat web data explorer.

Anchored poverty measures begin with the same fully or weakly relative measure in 1 year (t) and then compare relative poverty in some future year (say year $t+10$) with poverty measures against a poverty line that has been changed only for prices between year t and year $t+10$. These measures are especially useful in periods of rapid expansion or contraction in an economy, where relative poverty may not change by a lot, but where absolute poverty does change because of economic growth or contraction (see Atkinson et al., 2002; Johnson and Smeeding, 2012; OECD, 2013; Smeeding, 2006). Any absolute poverty line is also, therefore, an anchored poverty line. The difference is that an anchored poverty line can be updated to any period that is relevant to policy, given the analysis. As suggested above, the absolute (or anchored) U.S. Orshansky poverty line for the 1960s was about the same as a fully relative half-median income measure at that time. The United States has anchored its "official" poverty measure at this same point since that date. Now, 60 years later, the U.S. poverty line is only at about 30% of median income, not the 50% it was at its inception. Hence, analysts prefer to anchor their U.S. poverty studies at a semi-relative line (Johnson and Smeeding, 2012).

Here, for simplicity and breadth, we focus exclusively on the "headcount" measure of poverty, the share of people who fall below some definable point that indexes poverty. This approach does not measure the depth of economic need, the poverty gap, or the severity of poverty. People who are poor could become richer or poorer, with no change in a headcount measure of poverty. A pragmatic reason for using the poverty gap is that the headcount may be quite sensitive where there are spikes in the distribution because of the payment of flat-rate social transfers such as minimum social retirement level (or changes to the minimum wage).[14] Others (see especially Foster et al., 1984; Sen, 1976) focus on poverty measures that examine the distribution of poverty among the poor, taking account of both the depth of poverty and its severity. Because headcount measures are more easily understood, compared, and implemented than other more complex measures, we rely on them below.

The data we use are taken from LIS and sometimes OECD and are limited mainly to rich countries and MICs. The OECD includes a large number of rich nations, as well as Chile, Mexico, and Turkey. Both LIS and OECD have been interested in the BRICS countries (Brazil, Russia, India, China, and South Africa). LIS also has expanded to include other Latin American nations and Mexico. To establish trends in income poverty, however, one must have at least a decade or two of data, and here the number of MICs we can examine is severely limited.[15]

[14] A pragmatic reason for not using the poverty gap, especially in cross-national studies, is that underreporting of incomes, the definition of incomes, and editing for item nonresponse may differentially affect the lowest incomes and overstate the poverty gap.

[15] Indeed, we do not use the Eurostat (2012) poverty measures for two reasons. First, both the LIS and OECD measures rely on the same EU-SILC data for most of the EU nations and, second, because the EU-SILC data are very recent, starting only in 2005.

8.3.4 Level of and Trends in Poverty

We examine the level of and trends in poverty in a set of graphs and one table, all based on the LIS key figures data set, plus some special tabulations, to determine the level of anchored poverty using both LIS and OECD data. The percentage of people living below the half-median poverty line can now be examined for 38 nations using the LIS data (Figure 8.2). The 28 nations with light gray bars are the richest Anglo-Saxon, EU, and OECD nations; the 10 darker gray bars are for the MICs, including Russia, the BRICS nations, and several South American nations.[16]

If a "less poor" country is one with a single-digit poverty rate (where between 5% and 10% of its population are poor), 17 countries hit that target in the mid- to late 2000s, as

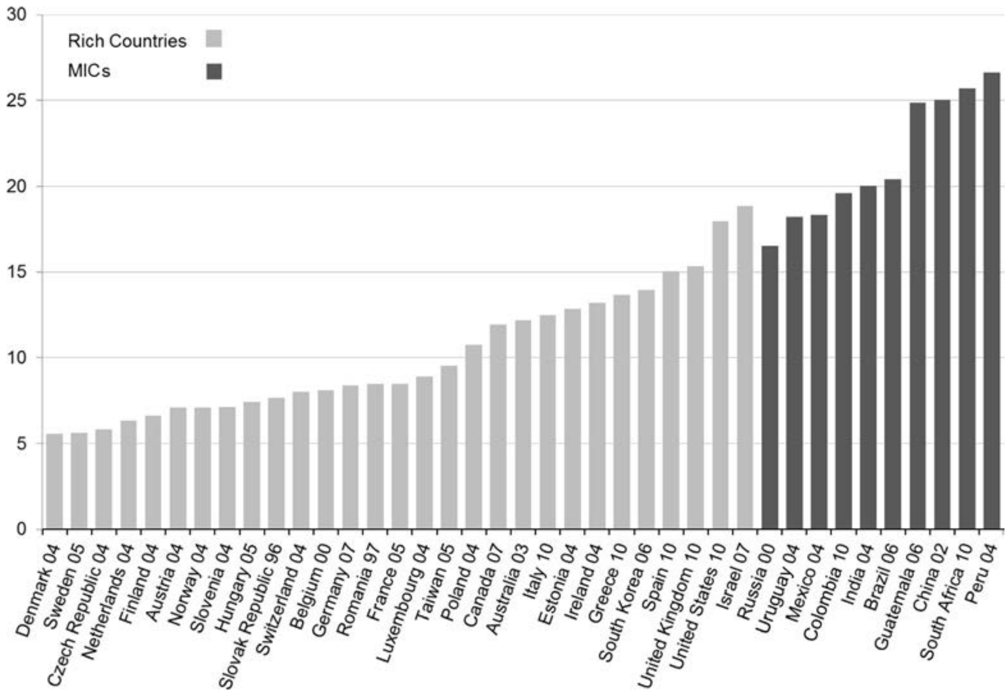


Figure 8.2 Relative poverty rates for total population (mid- to late 2000s) using LIS data. Poverty is measured by the percentage of people living in households with income (adjusted for family size) below half the median national income. *Source: Luxembourg Income Study, Key Figures: http://www.lisdatacenter.org/lis-ikf-webapp/app/search-ikf-figures*.

[16] The Eurostat (2005) produces poverty measures for all 27 EU nations now, including some which are not captured in either OECD or LIS data, and measures of poverty depth and severity as well. But their figures are at the 60% of poverty level and are not comparable to the half median figures in LIS and OECD.

shown in Figure 8.2. The Scandinavian and Nordic nations are generally lowest, along with a number of "middle" western, central, and eastern European nations who have joined the 27 in the EU (from Belgium and the Netherlands west to Luxembourg, Germany, France, and Austria, plus Switzerland, the Czech Republic, Slovakia, Hungary, Slovenia, and Romania). This pattern has been more or less the same since the first LIS measures appeared 20–25 years ago (Atkinson et al., 1995; Smeeding et al., 1990), although the number of nations has now expanded considerably. Taiwan weighs in with the 17th lowest poverty rate—about 9.5%. Another nine nations have relative poverty rates from 10% to 15%, including Italy, Spain, Greece, Poland, Estonia, Canada, Australia, Ireland, and South Korea. Three rich nations are between 15% and 19%: the United Kingdom (15%), the United States (18%), and Israel (19%). Moving to the MICs, six countries overlap the three rich nations in the 15–20% range; Russia has a poverty rate below the United States and Israel, and Uruguay and Mexico more or less even with the United States. Finally, Colombia, India, and Brazil were all at 20% poverty. Poverty rates are 25% and above in Guatemala, China, South Africa, and Peru. In short, the range of comparable relative poverty rates from the most comparable source extant varies by a factor of 5.

The OECD data in Figure 8.3 provide essentially the same picture but measuring all nations in 2010 compared with 2002–2010, as shown in Figure 8.2. The OECD data also

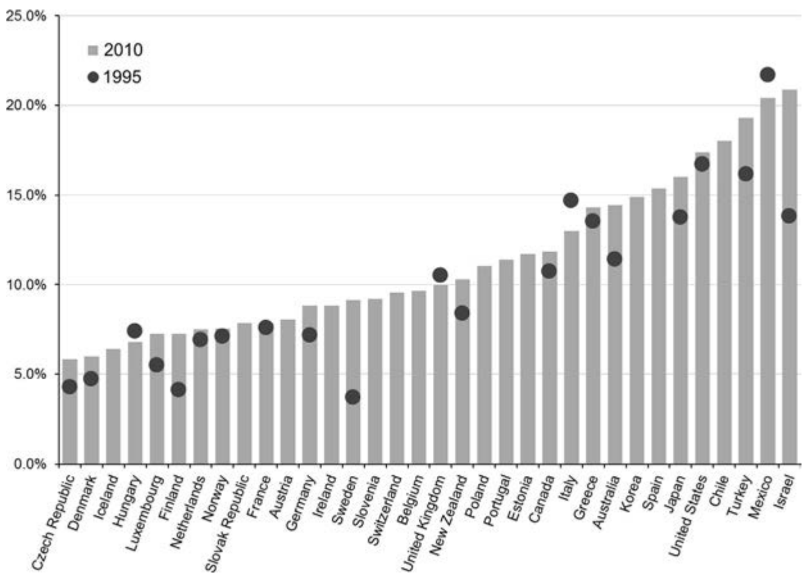


Figure 8.3 Levels and trends in relative poverty in OECD nations: 1995–2010. Poverty is measured by the percentage of people living in households with income (adjusted for family size) below half the median national income. *Source: OECD Income Distribution Database (www.oecd.org/social/income-distribution-database.htm).*

add a few nations (Iceland, Chile, and Turkey) to those in Figure 8.2 and also presents some data on 15-year trends in poverty, where available. Here, Israel leads the league in the table of poverty, with headcount rates surpassing 20%. The advantage of the OECD data is its rapidity of observation, and with 15-year trends, it is clear that relative poverty rates may change substantially over short periods of time.

Poverty in LIS is typically somewhat higher among children (Figure 8.4). Poverty averaged 13.5% among the countries for the total population but 16.5% for children. The correlation between child poverty and poverty in the total population is, however, quite high at 0.91, as reflected in Figure 8.4. The slope of the regression line in Figure 8.4 is 1.32, suggesting that child poverty rises about one-third faster than does overall poverty in these nations. The same sets of countries that are high, middle, and low poverty countries in Figures 8.2 and 8.3 also fall in the same relative positions for child poverty, but in some nations, such as Uruguay and Brazil, child poverty is disproportionately higher than overall poverty. In South Korea, child poverty is substantially lower than overall

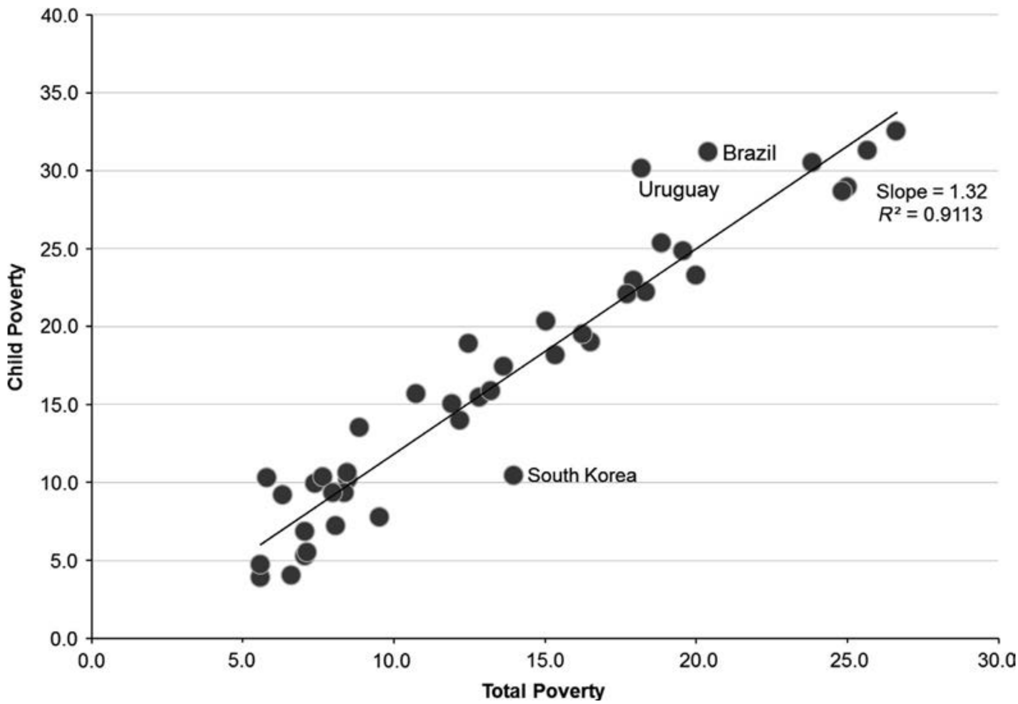


Figure 8.4 Correlation between total population poverty and child poverty in 38 rich and middle-income countries (late 2000s) using LIS data. Poverty is measured by the percentage of people living in households with income (adjusted for family size) below half the median national income. *Source: Luxembourg Income Study, Key Figures: http://www.lisdatacenter.org/lis-ikf-webapp/app/search-ikf-figures.*

poverty. In the others, child poverty and overall poverty track each other closely. In general, poverty among the elderly is both lower and falling compared with that among children, which is higher and rising in most nations (LIS key figures, and OECD, 2013, Figure 8).

Trends in poverty can be evaluated using the same data and allow us to break the countries into several different groups based the range of years over which data are available and geographic/institutional comparability. The panels in Figure 8.5 show longer-term (since 1979) trends for the 14 different countries that have been in the LIS data for the longest period. All of these figures include data up through 2010 (or the latest year available). The trends from 1995 to 2010 are best illustrated using the OECD data in Figure 8.3, where we have such data for 21 nations.

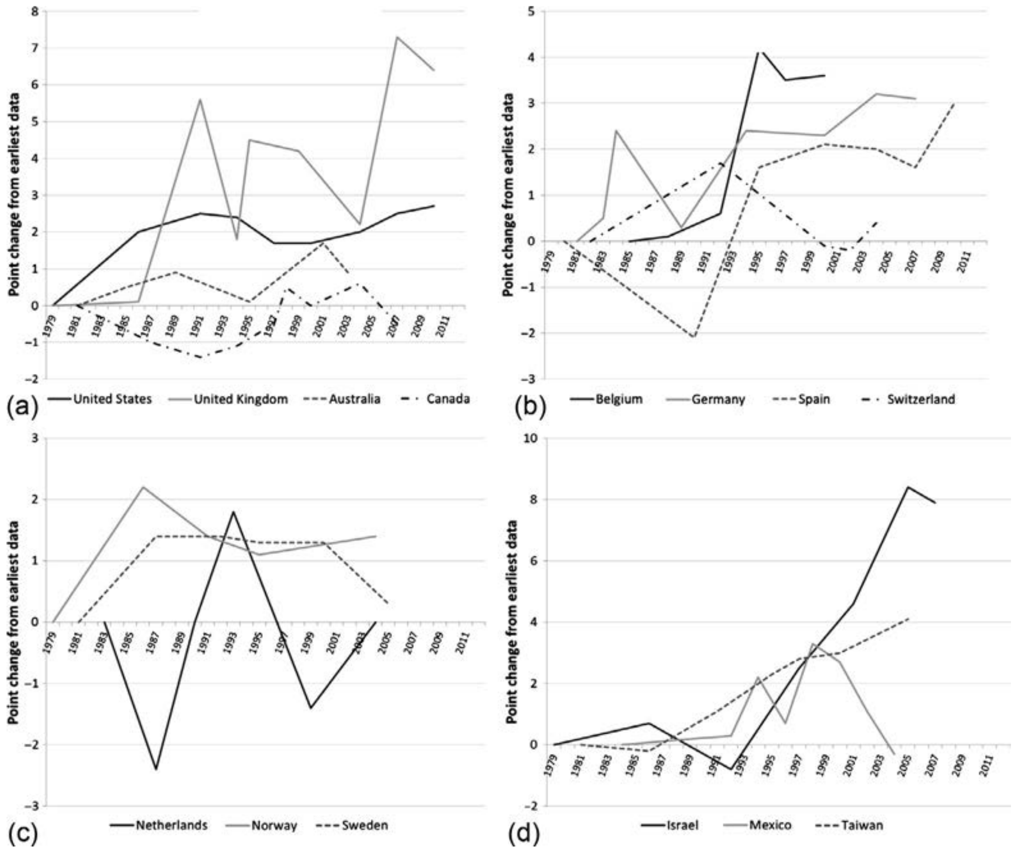


Figure 8.5 Trends in half-median poverty for 14 rich/middle-income nations, 1979–2010: (a) English-speaking countries, (b) continental/southern Europe, (c) Nordic countries, and (d) other countries. Poverty is measured by the percentage of people living in households with income (adjusted for family size) below half the median national income.

In analyzing trends in poverty, we are interested in both the direction of change and its magnitude. One finding is that none of the countries in Figure 8.5 (those countries with the longest series of data) have poverty that is appreciably (3 percentage points) lower in the most recent year than in the initial year of data from the late 1970s or early 1980s.[17] Canada (panel a) and Mexico (panel d) do have a bit lower poverty, but the difference for each country is very small (−0.5 and −0.3 points, respectively, with both series extending up through the mid-2000s—2007 and 2004, respectively). The trend data from the OECD (1995–2010 in Figure 8.3) suggest that relative poverty decreased only in Italy and Mexico over that period, although both by less than 3 percentage points.

Each of the other countries with long trends in Figure 8.5 has seen poverty increase or remain flat. Two countries stand out for particularly large increases, including Israel (panel d) and the United Kingdom (panel a), whereas the Nordic countries stand out as a cluster for seeing very little change in poverty (panel c) based on the LIS data. In contrast, the OECD data in Figure 8.3 show a massive increase in Swedish poverty, coming mostly after 2005 (compare with the LIS trend for Sweden in Figure 8.5c) and almost no change from 1995 to 2010 in the United Kingdom.[18] While relative poverty more than doubled in Sweden, appreciable increases can also be found in Australia, Finland, Israel, and Turkey over the 1995–2010 period (Figure 8.3).

Returning to Figure 8.5, in some nations, such as the Netherlands and Spain, poverty fell in the 1980s but returned to former levels (Netherlands) or went on to new heights (Spain) based on the LIS data. Poverty in Mexico rose though 1997, but then plummeted back to near its origin in 2004. Israel and the United Kingdom each had poverty rates more than 6 points above their origins by the late 2000s. Poverty rose steadily in the United States and Germany, increasing by about 3 percentage points in each, and by 4 points in Taiwan and more than 3 points in Belgium, from origin until the late 2000s. The rest generally stayed within ±3 percentage point bands from the origin until the final year.

While lessons about the importance of the start and end dates in term of volatility can be drawn, as well as differences across data sources in Figures 8.3 and 8.4, some other lessons emerge. These trends suggest that progress against relative poverty was uneven

[17] *Appreciably* here means more than a 3 percentage point change. Atkinson and Morelli (2012) discuss the definition of a salient change in the poverty percentage, explaining that there are both supply (sampling error and other design elements) and demand considerations (use of the figures). They end up applying a 2 percentage point change criterion. The period examined here is a much longer one, so we choose 3 percentage points as the cutoff. The lines in Figure 8.5 show the 3 percentage point bounds in each panel.

[18] Interestingly, the increase in Swedish relative poverty coincides with a rapid increase in income inequality (see Section 8.3.5). Because Sweden is one of the few nations that use register data that contain the full range of top incomes, the rapid increase in the Swedish median income may in part be driving this trend.

Table 8.1 Trends in relative and anchored poverty

		Poverty rates				
		Initial year	End year		Percentage point change from initial year	
	Years	Relative	Relative	Anchored	Relative	Anchored
Czech Republic	1996–2004	5.1	5.8	3.4	0.7	−1.7
Germany	1994–2007	7.7	8.4	7.3	0.7	−0.4
France	1994–2005	8.0	8.5	7.2	0.5	−0.8
Netherlands	1993–2004	8.1	6.3	4.4	−1.8	−3.7
Hungary	1994–2005	9.9	7.4	4.8	−2.5	−5.1
United Kingdom	**1994–2010**	**10.8**	**15.4**	**7.2**	**4.6**	**−3.6**
Canada	1994–2007	11.3	11.9	7.6	0.6	−3.7
Australia	1995–2003	11.4	12.2	7.8	0.8	−3.6
Italy	**1995–2010**	**14.1**	**12.5**	**9.5**	**−1.6**	**−4.6**
Greece	**1995–2010**	**15.4**	**13.6**	**6.4**	**−1.8**	**−9.0**
United States	**1994–2010**	**17.6**	**17.9**	**14.5**	**0.3**	**−3.1**
Mexico	1994–2004	20.8	18.3	16.5	−2.5	−4.3
Average		11.7	11.5	8.0	−0.2	−3.6

Data are based on the authors' calculations from LIS microdata files, http://www.lisdatacenter.org/.
Note: Poverty is measured by the percentage of people living in households with income (adjusted for family size) below half of the median national income.

and rare in rich nations over the past 20–30 years. Other than Mexico, poverty rates did not consistently decrease over the past 25 years in any of the nations we examine here.[19]

8.3.5 Relative Versus Anchored Poverty and the GR

A different way to examine progress against poverty is to take a set of OECD nations and examine changes in both relative and anchored poverty in 12 nations over an 8–15-year period using LIS (Table 8.1) or across the shorter period of the GR, from 2005 or 2007 to 2010 (Figure 8.6). On average, relative poverty did not change much in the LIS, but anchored poverty fell by about a third from 11.7% to 8.0% between the mid-1990s and the year of the most recent observation (Table 8.1), suggesting rising living standards for people with incomes that would have been considered poor in the initial period. Indeed, anchored poverty decreased in every nation, reflecting rising living standards in Europe and elsewhere in the rich countries and MICs up until the GR. In contrast,

[19] Ferreira de Souza (2012) also suggests that both poverty and inequality decreased in Brazil over the 1995–2009 period.

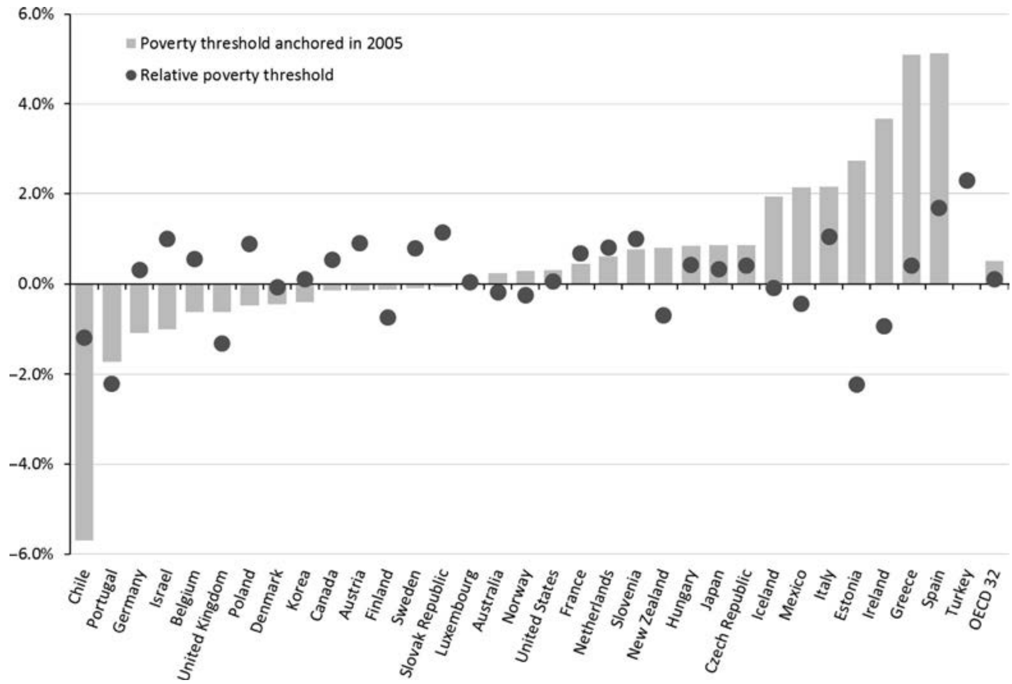


Figure 8.6 Anchored poverty in OECD countries: 2007–2010. *Source: OECD Income Distribution Database (www.oecd.org/social/income-distribution-database.htm).*

the changes in relative poverty over this same period were small, on average, in the LIS data but ranged from an increase of 4.6 percentage points in the United Kingdom to a decrease of 2.5 percentage points in Hungary and Mexico. All other relative poverty rates changed by less than 2 percentage points over this period.

The effects of the GR are included in the four LIS data sets in bold for the United States, United Kingdom, Italy, and Greece in Table 8.1. In each nation a data point is also available for 2007 (or 2008 for Italy only). In each nation, relative poverty rose by 0.2–2.2 percentage points through 2010, suggesting greater relative income losses for the poor than the rich in each nation during the GR. Despite the overall trends in each nation, anchored poverty increased between 2007/2008 and 2010. It increased by 1.2 points in the United States, 1.9 points in Italy, 2.6 points in Greece, and 3.0 percentage points in the United Kingdom. Hence, in each nation, despite the overall reductions in anchored poverty shown in Table 8.1, the poor lost ground in both relative and real terms over the course of the GR.

The OECD data (Figure 8.6) suggest much the same pattern in these four nations but add many others as well. Iceland, Mexico, Spain, Estonia, and Ireland join the list above, where living standards fell during the GR and anchored poverty increased much faster than relative poverty. Indeed, relative poverty did not increase much at all during the GR (and even decreased in Estonia and Ireland). In Poland, Belgium, and Germany, anchored poverty decreased but relative poverty did not change much. In Portugal and Chile, both anchored and relative poverty decreased during the GR. The changes in other nations were smaller.

We conclude that there was little progress in reducing relative poverty in almost all the rich nations examined here over the past two or three decades. Anchored poverty did decline in almost all rich nations from the 1990s up until the GR in 2007. Since the onset of the GR, however, anchored poverty has trended upward, with increases in anchored poverty in a majority of nations reducing some of the progress in real living standards for low-income households over the past 20 years, especially in the nations hardest hit by the GR. Relative poverty rates changed much less during the GR.

Finally, Figure 8.7 shows the correlation between relative poverty and inequality (using the LIS project Gini coefficient for DHI) across 38 nations. The correlation is astoundingly high: over 91%. The slope is 0.63, suggesting that a 10-point difference in the Gini, say from 0.20 to 0.30, is associated with a more than 6 percentage point increase in relative poverty. Still, at inequality levels of about 0.27 and 0.32, poverty rates can vary as much as 4 percentage points across nations with the same level of overall

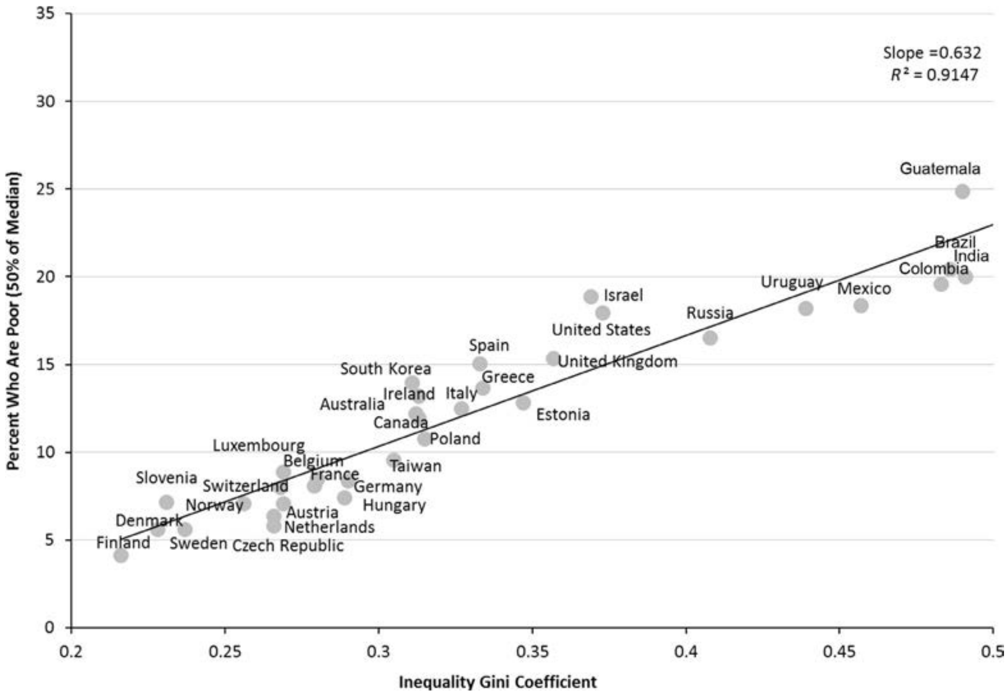


Figure 8.7 Relationship of relative poverty and income inequality in 38 nations (LIS). *Source: Luxembourg Income Study, Key Figures: http://www.lisdatacenter.org/lis-ikf-webapp/app/search-ikf-figures.*

inequality. Four nations stand out as having above-average poverty for their inequality level: Guatemala, Israel, South Korea, and the United States. Relative poverty levels are notably lower than inequality in the Czech Republic, Netherlands, Austria, and Hungary.

8.4. INEQUALITY IN INCOME

8.4.1 Measures of Inequality from the Overall Distribution

8.4.1.1 Introduction

This section focuses on measures of the overall distribution of income in high-income and some middle-income and developing countries. In contrast with the next section, which focuses narrowly on the top of the pretax income distribution, this section considers a variety of statistics that either explicitly exclude the very top (and bottom) of the distribution or that use the full distribution but are calculated with data that are not necessarily representative of incomes at the very top. Most of this section describes trends since 1970, but some attention is also paid to data series that are available over shorter periods and to a single-year analysis of the most current available income data, which allow us to discuss a broader range of inequality metrics and a greater number of countries.

Overall conclusions about the broad distribution of household income include:

* The countries with the least unequal distributions are the Nordic (Sweden, Norway, Denmark, and Finland) and "Benelux" (Belgium, Netherlands, and Luxembourg) countries as well as Austria and some eastern European nations.
* Across MICs and high-income countries there is a wide range in levels of inequality. By most measures the income distribution in the United States is among the most unequal, and when compared with the narrower set of the richest nations, the distribution in the United States is the most unequal. A number of MICs and developing nations, including Brazil, China, Turkey, and South Africa, though, have income distributions that are more unequal than in the United States.
* Taxes and transfers reduce the degree of inequality in every country, but there is dramatic variation in the extent of redistribution. The impact of taxes and transfers is very small in some highly unequal countries (Russia) and some less unequal ones (South Korea). In some countries, taxes and transfers have a dramatic impact on the distribution of income; Finland has among the most unequal distributions of market income but one of the most equal distributions of DHI because of the extensive distribution in its welfare state. The United States combines relatively high levels of inequality in market income with very low levels of tax and transfer redistribution to achieve the highest level of DHI inequality among rich nations.
* The distribution of income has become more unequal in most countries since the 1970s. The only rich country to buck the long-term trends toward greater

inequality is France. Even France, though, has experienced increases in inequality since the early 2000s.

* The income distribution in a number of countries has followed a U-shaped pattern (Sweden, Finland, and Canada), falling in the 1970s or the 1980s before rising in the 1990s.
* Two of the most unequal of the rich nations—the United States and the United Kingdom—experienced large increases in inequality in the late 1970s and 1980s and modest increases in the second half of the 1990s, but in both countries the level of inequality in 2010 was not very different from levels experienced in the early 1990s.
* The distribution of market income in Germany, Italy, Japan, and some of the Nordic countries grew steadily more unequal between the mid-1980s and the mid-2000s, and the distribution of pretax/transfer income in those countries is now almost as unequal as in the United States, Israel, or the United Kingdom.
* In almost all countries the long-term trends in inequality are more pronounced among the working-age population.

8.4.1.2 Distributional Statistics

A variety of statistics have been developed for the analysis of the distribution of income. The most commonly used statistic is the Gini coefficient, but a number of other measures have been applied to a wide range of countries using data covering the most recent decades. The statistics discussed below include Lorenz curves, the Gini coefficient, Atkinson Index (ATK), percentile ratios (P90/P50 and P90/P10), quintile shares (S80/S20), and the Palma Index. (See Allison, 1978; Atkinson, 1970; Cowell, 2000; Heshmati, 2004; among others, for overviews of the various summary statistics to describe distributional inequality.)

Not a statistic *per se*, the Lorenz curve is a graphical representation of the cumulative distribution of income. The Lorenz curve uses ordered income data and shows the cumulative share of income held at each point in the distribution of households.

To reduce the information contained in the Lorenz curve to a single number, a variety of summary statistics have been proposed. One that has a direct link to the Lorenz curve is the Gini coefficient. The Gini coefficient can be calculated in a number of ways and visually can be represented as a ratio of the area between the Lorenz curve and the perfect equality line divided by the total area below the perfect equality line. In ordered data for household share of total income, the 45-degree line represents perfect equality; each household has the same income and each point in the distribution of total households matches the same point in the distribution of total household income (e.g., the bottom 45% of households receive 45% of total income). The Gini coefficient ranges from 0 (perfect equality) to 1 (the most extreme inequality) if all income is held by a single household.

Using unordered data, the Gini coefficient for household income can be calculated as the relative mean difference, or the average absolute difference between incomes for all pairs of households divided by twice the mean income (Allison, 1978):

$$\text{Gini} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\left|x_i - x_j\right|}{2N^2\overline{x}}, \tag{8.1}$$

where N represents the total number of households, i and j index each household in all possible pairings of household, x is household income, and $\overline{x}$ is mean income over the sample.

The Gini coefficient is one of many statistics representing the entire distribution. Other commonly used measures of inequality focus on specific points or regions of the distribution. Below we discuss inequality measures from the most recently available data using the P90/P10 and P90/P50 interdecile ratios, which represent "high" income levels (from the 90th percentile of the distribution in this case) as some multiple of "low" income (the 10th percentile of the distribution) or "middle" income (the median). A similar measure, the S80/S20, represents a ratio of the shares of total household income received by those in the top quintile of the distribution and those in the bottom quintile. The Palma Index, popularized by Palma (2011), is a slight modification of the more common S80/S20 and divides the share of income held by the highest 10% of the distribution by the share of income received by the lowest-income 40% of the distribution.

The final measure discussed in this section is the ATK.[20] Similar to the Gini coefficient, the ATK summarizes the entire distribution. Unlike the Gini, though, the ATK can be decomposed to identify different groups or income sources making different contributions to inequality. The ATK differs from the previous measures by explicitly incorporating a weighting variable that can be selected to place more weight on incomes at the top or the bottom of the distribution

$$\text{ATK}(\varepsilon) = 1 - \frac{1}{x}\left(\frac{1}{N}\sum_{i=1}^{N} x_i^{1-\varepsilon}\right)^{\frac{1}{(1-\varepsilon)}}. \tag{8.2}$$

The weighting variable (ε) is typically selected from values between 0 and 2, although any positive value can be used. Higher values for the weight, called a measure of inequality

[20] The mean log deviation (MLD) is another statistic that uses the entire distribution but tends to produce results very similar to the Gini coefficient. MLD statistics are not included here because of limited space, but they have been calculated by the OECD in the past, including in Divided We Stand (2011). Also, the squared coefficient of variation (SCV) has been used in some analyses of income distribution, including OECD (2011), but rankings developed using this measure are very sensitive. Deding and Dall Schmidt (2002) showed that, compared to the Gini coefficient, the SCV produces substantially larger year-to-year shifts in inequality and is particularly sensitive to tax and transfer payments at the upper tail of the distribution. For these reasons we do not include SCV measures in this review.

aversion, reflect greater sensitivity to incomes at the lower end of the distribution. The ATK falls between 0 and 1, equaling 0 under perfect equality and with higher values when dispersion is greater.

8.4.1.3 Levels of Inequality in High- and Middle-Income Countries in the Late 2000s

With expanded interest in the distribution of income, there are more data available from recent years to compare incomes across countries than at any point in history. This section reviews evidence from a broad array of rich countries and MICs using all of the distribution statistics described above. The following section focuses on a narrower set of countries and examines trends in the distribution of income using a more limited set of statistics. All of the analyses in these sections rely heavily on the data produced and made available by LIS, Eurostat, the OECD, and the national statistical agencies of a handful of rich countries.

8.4.1.3.1 Lorenz Curves

Unlike most summary statistics used in the analysis of inequality, Lorenz curves visually represent the entire distribution. Analyzing these plotted cumulative distribution functions allows us to see whether pairs of countries can be ranked by standard dominance criteria.[21] Figure 8.8 includes a series of Lorenz curves for different geographically or institutionally coherent clusters of countries. Each graph also includes the Lorenz curve for the United States to aid in comparability across the different graphs. The figure uses data from the most recent LIS wave for each country (identified in the individual graphs) and represents equivalized DHI.[22]

The distribution of income in the continental European countries (including Austria, Belgium, France, Germany, Luxembourg, and Switzerland), as well as Japan (shown in Figure 8.8a), and the Nordic countries (Denmark, Finland, Norway, and Sweden, shown in Figure 8.8b) is much less unequal than in the United States. Because the Lorenz curves do not cross at any point, we can say that each of these countries has a "superior" Lorenz curve to the United States. Any differences between these countries—which are slightly more evident among the Nordic counties—are small compared with their differences with the United States.

The U.S. distribution is more unequal than most of the rest of the European countries, but not to such a great extent. In the case of the Anglo-Saxon countries (Figure 8.8c), Australia, Canada, and Ireland have Lorenz curves that are superior to that of the United States, but the Lorenz curves for the United States and the

[21] See Cowell, 2000 (*The Handbook of Income Distribution*, Chapter 1) for a discussion of the properties of Lorenz curves and dominance criteria.

[22] The Lorenz curves displayed here are based on income that is bottom coded at 1% of average income and top coded at 10 times the median household income.

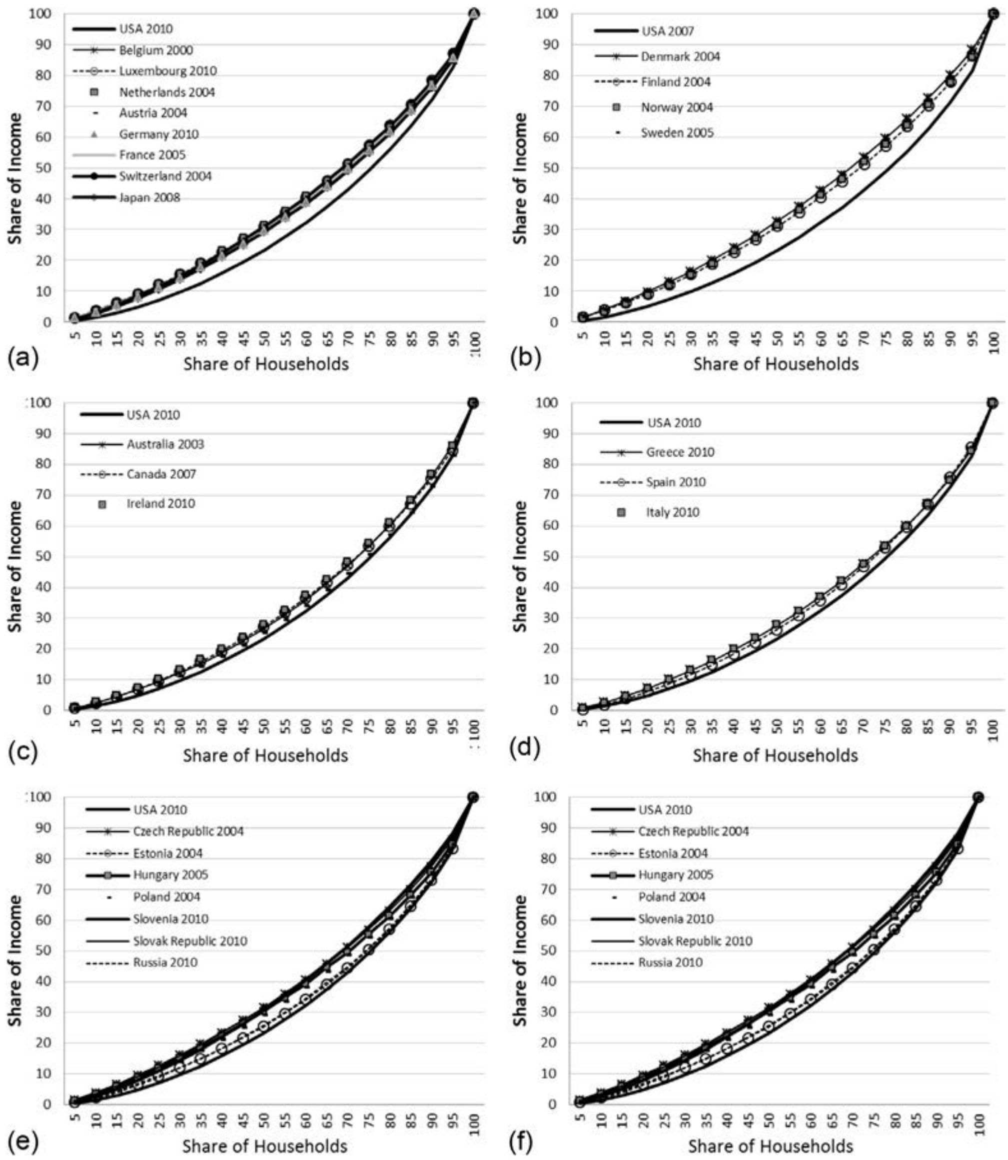


Figure 8.8 Lorenz curves of equivalized DHI (LIS) in the mid- and late 2000s: (a) continental European countries (and Japan), (b) Nordic countries, (c) Anglo-Saxon countries, (d) Southern European countries, (e) Eastern European countries, and (f) other countries. *Data are based on the authors' analysis of LIS data.*

United Kingdom are virtually indistinguishable, although they do not cross. The United States also has an inferior Lorenz curve relative to the countries in Southern Europe (Spain, Italy, and Greece, shown in Figure 8.8d), but the gaps are less dramatic than for the Nordic or continental European countries. None of the southern European countries has a distribution that is superior to the others, as the Lorenz curves cross at the top and the bottom of the distributions.

Even in Eastern Europe (Figure 8.8e), each country has a Lorenz curve superior to that of the United States. In the case of Estonia (2004) and the Russian Federation, the distribution is very similar, especially in the upper third, but at no point do the Lorenz curves cross. In the Slovak Republic, Slovenia, and the Czech Republic (2004) the distributions look more similar to those of continental European countries than those of their Eastern European neighbors.

Only when we expand the set of countries beyond Europe and include MICs and developing counties do we find distributions of income that are more unequal than that of the United States ("Other Countries" include South Korea, India, China, Brazil, and Israel, shown in Figure 8.8f). The most recent LIS data for Brazil, India, China, and South Africa show that the Lorenz curves for those countries are inferior to that of the United States. Among those four nations, South Africa stands out with the most unequal distribution. Israel and the United States have virtually indistinguishable Lorenz curves, both of which are inferior to the Lorenz curves for South Korea and Taiwan.

8.4.1.3.2 EU and OECD Country Summary Statistics and Rankings

In recent years the EU and the OECD have calculated timely summary distributional statistics for their member countries. These figures are based on DHI data from 2010 to 2011 for the EU and "around 2010" for the non-EU OECD countries.[23] Statistics from both entities are adjusted for household size using slightly different equivalence scales.

Figure 8.9 includes three different summary statistics for the 23 richest nations that are EU or OECD members and is sorted based on rankings for the Gini coefficient (shown in Figure 8.9a).[24] With a Gini coefficient of 0.38, the United States has the highest level of inequality among the rich nations. At the other extreme, with Gini coefficients between 0.23 and 0.28, the Nordic and Benelux (Belgium, Netherlands, and Luxembourg) countries and Austria had the most equal distribution of income, led by Norway. The large continental economies and the Anglo-Saxon countries fall in the middle, with Gini coefficients ranging from 0.28 and 0.31 in Germany and France, respectively, to between 0.32 and 0.33 in Australia, Canada, and the United Kingdom.

While they are based on smaller ranges of the distribution, the S80/S20 interquartile share ratio (Figure 8.9b) and the P90/P10 interdecile ratio (Figure 8.9c) each produce rankings similar to that of the Gini coefficient. In the rich nations with the highest

[23] All figures are reported based on the income year, not the survey year. Eurostat figures are originally published by survey year, but we report them according to the year in which the income was received. Details on the timing for the data are contained in Table 8.2 and Figure 8.9.

[24] These nations all have gross domestic product per capita of more than $29,000 (PPP) in 2012 and account for 23 of the world's 31 richest countries (International Monetary Fund, 2013).

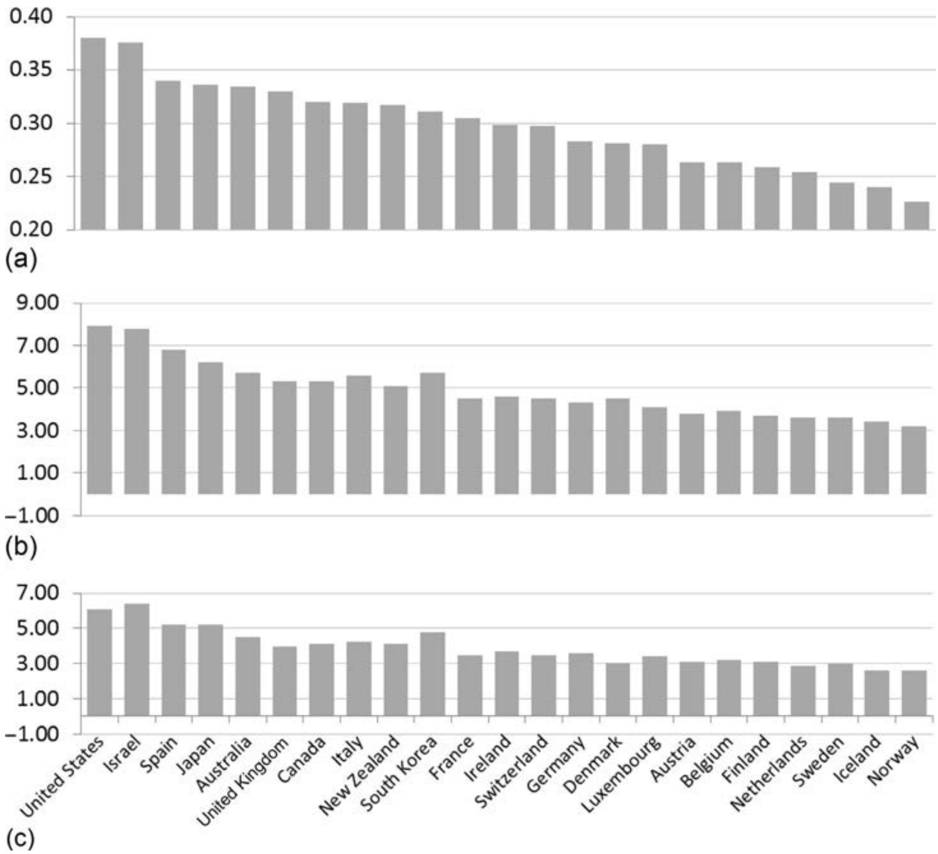


Figure 8.9 Summary distribution statistics for equivalized DHI for the richest EU and OECD nations for 2010–2011: (a) Gini coefficient, (b) interquartile share ratio (S80/S20), and (c) interdecile ratio (P90/P120). EU member country data are mainly from 2011 or 2010; non-EU OECD member country data are primarily from 2010. Gini coefficient, S80/S20 ratio, and P90/P10 ratio figures for EU member countries are based on Eurostat data and are mostly from 2010. A number of EU countries have data from 2011, including Denmark, Finland, France, Germany, Iceland, Luxembourg, the Netherlands, and Norway. Non-EU OECD member county figures for Gini, S80/S20, and P90/P10 are mainly from 2010, with some exceptions: South Korea, 2011; Japan, New Zealand, and Switzerland, 2009. *Sources: Eurostat and OECD.*

S80/S20 ratios, the United States and Israel, the average income among the highest-income fifth of households is 7.8 times the average income in the bottom fifth. In the less unequal Nordic and Benelux nations, the ratio ranged from 3.2 to 3.9. The P90/P10 ratio was 6.4 in Israel, followed closely by the United States at 6.1. Most of the ranking using the P90/P10 is similar to the S80/S20 ranking and, in turn, the Gini coefficient ranking, but the rich Asian nations stand out somewhat. In the Gini coefficient rankings, Japan and South Korea were similar to, and somewhat less unequal than,

Table 8.2 Summary distributional statistics for equivalized disposable household income—EU and OECD measures for 2010–2011 and the late 2000s

	Gini	Interquintile share ratio (S80/S20)	Interdecile ratio (P90/P10)
Australia	0.334	5.7	4.5
Austria	0.263	3.8	3.1
Belgium	0.263	3.9	3.2
Bulgaria	0.336	6.1	4.9
Canada	0.320	5.3	4.1
Croatia	0.31	5.4	4.5
Cyprus	0.31	4.7	3.7
Czech Republic	0.249	3.5	2.9
Denmark	0.281	4.5	3.0
Estonia	0.325	5.4	4.4
Finland	0.259	3.7	3.1
France	0.305	4.5	3.5
Germany	0.283	4.3	3.6
Greece	0.343	6.6	4.9
Hungary	0.269	4.0	3.3
Iceland	0.240	3.4	2.6
Ireland	0.298	4.6	3.7
Israel	0.376	7.8	6.4
Italy	0.319	5.6	4.2
Japan	0.336	6.2	5.2
Latvia	0.359	6.5	5.1
Lithuania	0.32	5.3	4.4
Luxembourg	0.280	4.1	3.4
Malta	0.272	3.9	3.3
Netherlands	0.254	3.6	2.9
New Zealand	0.317	5.1	4.1
Norway	0.226	3.2	2.6
Poland	0.309	4.9	4.0
Portugal	0.345	5.8	4.6
Romania	0.332	6.2	5.2
Russian Federation	0.428	9	6.9
Slovak Republic	0.257	3.8	3.1
Slovenia	0.237	3.4	3.0
South Korea	0.311	5.7	4.8
Spain	0.340	6.8	5.2
Sweden	0.244	3.6	3.0
Switzerland	0.297	4.5	3.5
Turkey	0.448	11.3	8.5
United Kingdom	0.330	5.3	4.0
United States	0.380	7.9	6.1

Sources: Eurostat and OECD.
Note: Eurostat data are used for EU countries that are also OECD members. EU member country data mainly from 2011 or 2010; non-EU OECD member country data are primarily from 2010. Gini Coefficient, S80/S20 ratios, and P90/P10 ratio figures for EU member countries are based on Eurostat data and are mostly from 2010. A number of EU countries have data from 2011, including Bulgaria, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Luxembourg, the Netherlands, Norway, Poland, Portugal, and Slovenia. Non-EU OECD member country figures for Gini, S80/S20 ratios, and P90/P10 ratios are mainly for 2010, with some exceptions: South Korea figures are from 2011; Japan, New Zealand, and Switzerland are from 2009; and Russian Federation are from 2008.
SCV for all countries is from the OECD's "Divided We Stand" and are primarily for 2008, except for Hungary and Turkey (2007) and Japan (2006). These statistics are no longer collected by the OECD.

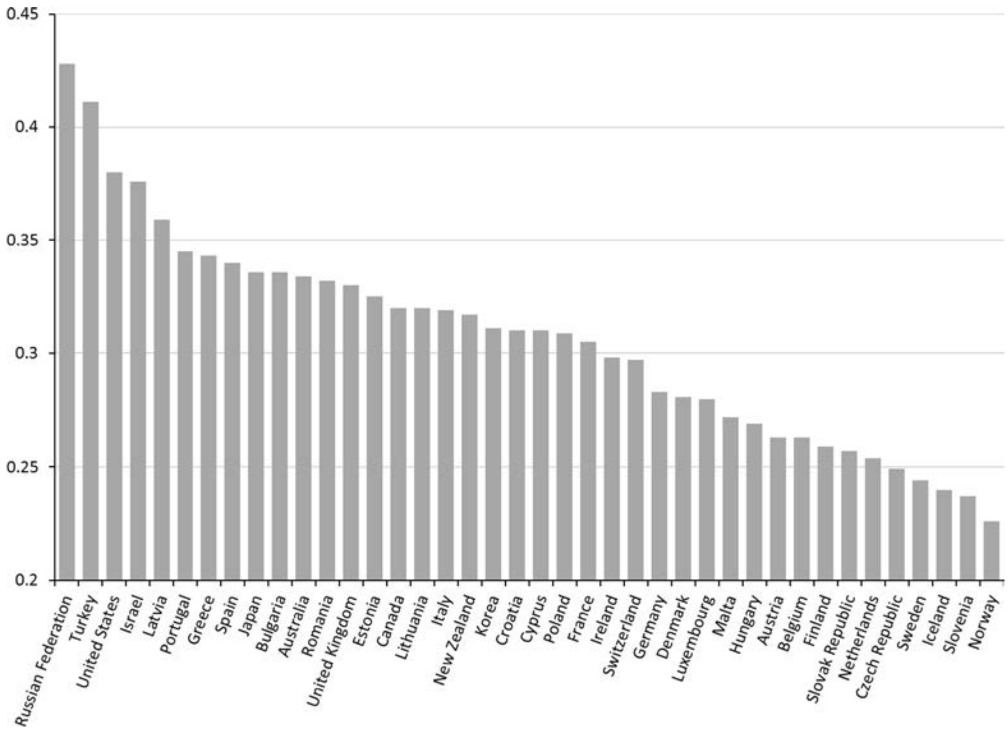


Figure 8.10 Gini coefficients for equivalized DHI for 2010–2011, including middle-income and developing EU and OECD nations. *Source: EU-SILC, Eurostat data (figures for 2010/2011 for EU member countries). OECD data are mainly from 2010; exceptions include South Korea (2011) and New Zealand and Turkey (2009).*

many of the Anglo-Saxon and southern European nations. Using the P90/P10, Japan and Korea appear more unequal and rank third and sixth, at 5.2 and 4.8, respectively. Among the less unequal Nordic and Benelux countries, the P90/P10 lies between 2.6 and 3.2.

The list of countries regarded as the "most unequal" or "least unequal" is, of course, somewhat dependent on the set of countries included. Figure 8.10 represents the ordered Gini coefficients for a set of countries that includes the 23 rich nations already shown in Figure 8.9 with 17 additional high-income countries, with gross domestic product per capita above $12,500 (International Monetary Fund (IMF), 2013), that are also part of EU or the OECD. In Figure 8.10, the United States is supplanted by the Russian Federation and Turkey as having the most unequal distributions using the Gini coefficient. The list of countries with less unequal distributions is similarly bolstered as the Nordic and Benelux countries are joined by several central European nations, including Slovenia and the Czech Republic. All of the summary statistics from Figures 8.9 and 8.10 are in included in Table 8.2.

8.4.1.3.3 LIS Country Summary Statistics and Rankings

As seen in the Lorenz curves above, the LIS project includes data from a number of countries that are not part of the EU or the OECD. LIS also regularly calculates several distribution statistics not typically reported by the OECD or the EU. Figure 8.11 includes two different ATK measures ($\varepsilon = 0.5$, 1), the P90/P50 interdecile ratio and the Palma Index (Figure 8.11d), for 34 countries with values reported in any of the three most recent LIS waves (covering the decade of the 2000s). The data used in these figures also are included in Table 8.3.

The alternative summary statistics in Figure 8.11 maintain the same basic rank ordering among the rich countries shown in Figure 8.9, with the least unequal distributions found in the Benelux and Nordic countries and the most unequal found in the United States, Israel, and the United Kingdom. Including the MICs and developing countries that are part of the LIS project, though, alters the ranking considerably. South Africa stands out as the most unequal country by far among the 34, with an ATK of 0.29, 38% higher than second-ranked China. Using a somewhat larger inequality aversion parameter ($\varepsilon = 1$) results in higher measured ATK numbers but by and large preserves the rank ordering across nations (Figure 8.11a). With greater sensitivity to incomes at the bottom of the distribution, the Czech Republic's rank (from most unequal to least unequal) falls three spots, and Switzerland's rises five spots, but overall our understanding of which countries have more- or less-equal distributions of income is essentially unchanged by modest changes in the inequality aversion parameter.

Analysis of the P90/P50 interdecile ratio (Figure 8.11c) demonstrates the dramatic differences in the distributions of the rich EU and OECD countries from those of the MICs and developing countries in the LIS project. Israel is the rich nation with the highest P90/P50, with an equivalized DHI at the 90th percentile 2.3 times that at the median. Four LIS lower-income countries (Brazil, China, India, and South Africa) have P90/P50 ratios at least 40% higher than Israel.

Proponents of adopting the Palma Index have argued that it isolates the portions of the income distribution that are most volatile over time and across countries (Cobham and Sumner, 2013). Compared to the Gini Index (and the ATK) the Palma Index is also transparent as to which portions of the distribution are determining the measure of inequality. This feature is shared by the P90/P10, P90/P50, and S80/S20 measures. Country rankings based on the Palma Index, calculated using LIS data, are very similar to those obtained using more common measures. The Nordic countries have the least unequal distributions, with values ranging from 0.98 (Norway) to 0.82 (Denmark), whereas South Africa has the most unequal, with a Palma Index of 7.8. The United States has the highest Palma Index (1.75) among rich nations.

Figure 8.11 indicates that the country rankings are similar across all four inequality measures. South Africa is the most unequal among the 34 nations in the LIS data using all of the measures, whereas Denmark is the least unequal. The United States ranks fifth most unequal using three of the measures and seventh most unequal using the other (P90/P50).

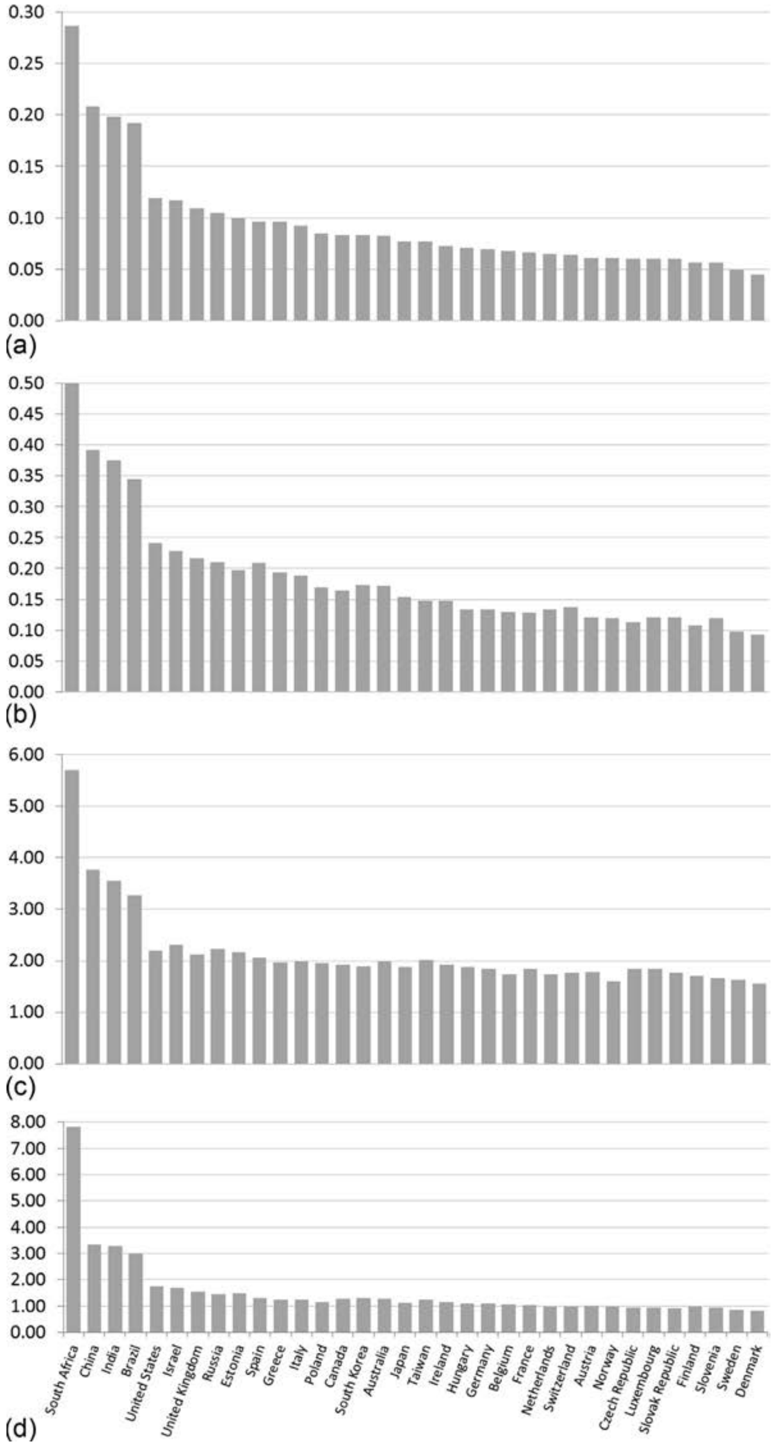


Figure 8.11 Distributional summary statistics from LIS countries using equivalized household income in the 2000s (LIS waves VI, VII, and VIII): (a) Atkinson coefficient ($\varepsilon = 0.5$), (b) Atkinson coefficient ($\varepsilon = 1$), (c) P90/P50 ratio, and (d) Palma Index (S10/S40 ratio). The sample years range from 2002 to 2010. *Data are from the authors' analysis of LIS data. (a)–(c) are from LIS published "key figures." (d) is based on the authors' analysis of LIS data.*

Table 8.3 Summary distribution statistics from LIS using equivalized disposable household income

	Atkinson coefficient ($\varepsilon = 0.5$)	Atkinson coefficient ($\varepsilon = 1$)	Percentile ratio (90/50)	Palma Index (S90/S40)
Australia, 2003	0.082	0.172	1.98	1.28
Austria, 2004	0.061	0.120	1.79	1.00
Belgium, 2000	0.068	0.129	1.74	1.08
Brazil, 2006	0.192	0.345	3.27	3.00
Canada, 2007	0.083	0.164	1.93	1.28
China, 2002	0.208	0.392	3.77	3.33
Czech Republic, 2004	0.060	0.113	1.85	0.96
Denmark, 2004	0.045	0.092	1.56	0.82
Estonia, 2004	0.100	0.197	2.17	1.49
Finland, 2004	0.056	0.108	1.71	0.98
France, 2005	0.066	0.128	1.84	1.04
Germany, 2010	0.069	0.133	1.85	1.10
Greece, 2010	0.096	0.194	1.97	1.26
Hungary, 2005	0.071	0.134	1.87	1.10
India, 2004	0.198	0.375	3.56	3.29
Ireland, 2010	0.072	0.147	1.92	1.14
Israel, 2010	0.117	0.228	2.30	1.69
Italy, 2010	0.092	0.189	1.99	1.26
Japan, 2008	0.077	0.154	1.88	1.13
Luxembourg, 2010	0.060	0.120	1.85	0.96
Netherlands, 2004	0.065	0.133	1.74	0.98
Norway, 2004	0.061	0.119	1.60	0.98
Poland, 2004	0.085	0.169	1.96	1.17
Russia, 2010	0.105	0.210	2.24	1.45
Slovak Republic, 2010	0.060	0.120	1.77	0.93
Slovenia, 2010	0.056	0.119	1.66	0.95
South Africa, 2010	0.287	0.505	5.70	7.81
South Korea	0.083	0.173	1.895	1.31
Spain, 2010	0.096	0.209	2.06	1.32
Sweden, 2005	0.049	0.097	1.63	0.85
Switzerland, 2004	0.064	0.137	1.76	0.97
Taiwan, 2005	0.077	0.147	2.02	1.26
United Kingdom, 2010	0.109	0.216	2.13	1.56
United States, 2010	0.119	0.241	2.19	1.75

Data are based on the authors' analysis of LIS project data. The Palma Index was calculated by the authors using LIS project data, the Atkinson coefficient, and P90/P50 from LIS published "key figures."

8.4.1.3.4 Comparing Current Distributions of Pretax and Transfer Income and DHI

In almost every nation, and particularly among rich nations, the tax and transfer systems reduce the disparity of income. Whether taxes are paid at higher rates among upper-income households, benefits and transfer payments are directed disproportionately toward lower-income households, or both, measures of inequality are lower for DHI than for market income. The extent to which the tax and transfer systems reduce measured inequality varies substantially across countries. The distributions of DHI and pretax and transfer income, and the extent to which taxes and transfers reduce inequality, are shown in Figure 8.12 for a set of 31 OECD countries. The figure shows Gini coefficients for DHI and pretax and transfer income (sorted on the latter) for all age levels (panel a) and for working-age (18–65 years old) individuals (panel b).

The rank ordering of countries based on inequality of pretax and transfer income for all ages (Figure 8.12a) is very different from the previously described rankings based on DHI. The United States does not have the most unequal distribution of pretax and transfer income, even among rich countries—it is ninth behind Ireland, Israel, the United Kingdom, and the southern European countries. The pretax and transfer income Gini for Italy is 0.50, 47% greater than South Korea, which has the lowest Gini among this set of countries. Also, instead of being clustered at the bottom, the Benelux countries are spread across the rankings based on the Gini coefficient for pretax and transfer income, and at least one Nordic country—Finland—rises to the middle.

Another important feature highlighted in Figure 8.12 is the substantial cross-national variation in the extent to which the tax and transfer systems reduce inequality. In several countries—notably the Russian Federation and South Korea—the tax and transfer systems have little impact on the distribution of income, and the Gini coefficient for DHI is only slightly smaller than the Gini for pretax and transfer income. In the case of Russia, low levels of redistribution leave the country with very high levels of inequality in DHI compared with other countries. In South Korea there is relatively little redistribution, but pretax and transfer income is distributed more evenly than in most countries, leaving a Gini for DHI that falls in the middle of the rankings.

In other countries, the tax and transfer system has a considerably larger impact on the distribution of income. In 11 countries the Gini coefficient is at least 40% lower for DHI than it is for market income. This is true for several of the Nordic and Benelux countries, as well as Ireland, Germany, and a number of eastern European countries. Substantial tax and transfer redistribution in these countries leaves them with the most equal distributions of DHI. In the case of the United States and Israel, above-average inequality in the distribution of market income combined with below-average levels of tax and transfer redistribution leave them with the highest Gini coefficients for DHI among the rich nations.

In this section we describe the difference between the Gini coefficients for pretax and transfer income and DHI as a measure of the extent of redistribution in a country. This measure of redistribution, however, has important limitations and warrants some caveats. One such caveat is that the gap between these two Gini coefficients is a distorted measure of "redistribution" because the tax and transfer policies carrying out said redistribution can be expected to cause some changes in household and firm economic behavior that

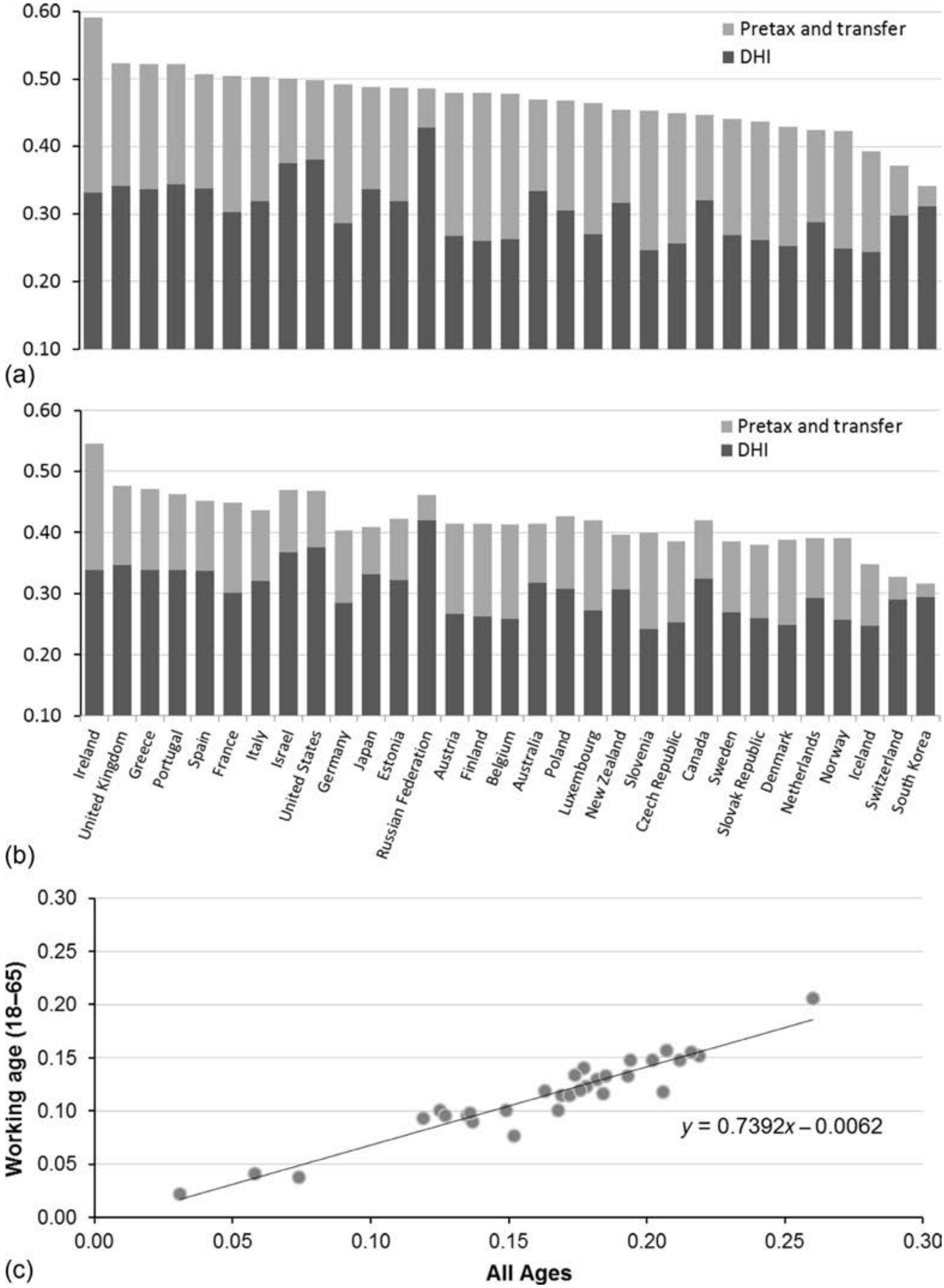


Figure 8.12 Gini coefficients around 2010 for pretax and transfer income and DHI for all ages (a), working age (18–65 years old) (b), and correlation between "redistribution" (pretax and transfer income Gini less DHI Gini) for all ages and the working-age populations (c). OECD member country data are primarily from 2010 with some exceptions: South Korea data are from 2011; Japan, New Zealand, and Switzerland data are from 2009; and the Russian Federation data are from 2008.

will be reflected in pretax and transfer income. Another limitation of this measure of redistribution is that similar types of income are classified as transfers in some countries but not in others, according to different institutional arrangements and policy choices. Retirement income systems are particularly relevant here. Countries with greater reliance on pensions provided directly by the public sector will seem to have greater redistribution than countries that finance retirement schemes through employers and private accounts (supported by tax incentives and potentially regulations).[25] A corollary is that countries with older populations (and otherwise equivalent pension systems) will seem to have greater redistribution by this measure.

We can compare the extent of redistribution across countries in a way that avoids some of these classification issues, at least in part, by using incomes from the working-age population (Figure 8.12b). Excluding retirees, who overwhelmingly rely on pension income, does not dramatically alter the rank ordering of countries based on the Gini coefficient for pretax and transfer income or the extent of redistribution observed across countries. The United States and the Anglo-Saxon and southern European countries remain the most unequal, whereas the Nordic and Benelux countries remain the least unequal. In a few countries, however, the cross-national ranking for inequality of pretax and transfer income jumps when elderly individuals are excluded; countries with notable increases include the United States, Canada, Israel, and the Russian Federation.

This extent of redistribution is greater among the total population than it is for the working-age population in every country. In the typical country the measure of redistribution for the working age is almost three quarters as large as it is for the total population (Figure 8.12c). The correlation between redistribution for all ages and for the working-age population is quite high. The simple correlation coefficient between the measures of redistribution for these two different age groups is 0.95. Countries that engage in relatively high levels of redistribution among the total population (including the elderly) also tend to engage in relatively high levels of redistribution among the working-age population. Table 8.4 contains all of the figures used in Figure 8.12.

8.4.1.4 Trends in the Distribution of Income Since 1970

Because income distribution data and statistics for some countries are only available for recent years, we are able to analyze trends in the distribution of income since the 1970s for a more limited set of countries than was discussed in the previous section. Here we first describe trends in the Gini coefficient for equivalized DHI since the mid-1970s for 10 rich nations. Then we turn to trends in the S80/S20 and P90/P10 measures, which are available for a somewhat larger number of OECD countries starting mostly in the

[25] This will be true even if the level of pension benefits is identical and the level of savings to finance the private pension scheme is equal to the taxes to finance the public pension scheme.

Table 8.4 Comparison of household market income and disposable household income: Gini coefficients for OECD countries around 2010

	All ages			Working-age population (18–65 years old)		
	Pretax and transfer income	**DHI**	**Reduction in Gini due to taxes and transfers**	**Pretax and transfer income**	**DHI**	**Reduction in Gini due to taxes and transfers**
Australia	0.469	0.334	0.135	0.414	0.318	0.096
Austria	0.479	0.267	0.212	0.414	0.266	0.148
Belgium	0.478	0.262	0.216	0.413	0.258	0.155
Canada	0.447	0.320	0.127	0.420	0.324	0.096
Czech Republic	0.449	0.256	0.193	0.386	0.253	0.133
Denmark	0.429	0.252	0.177	0.388	0.248	0.140
Estonia	0.487	0.319	0.168	0.423	0.322	0.101
Finland	0.479	0.260	0.219	0.415	0.263	0.152
France	0.505	0.303	0.202	0.449	0.301	0.148
Germany	0.492	0.286	0.206	0.403	0.285	0.118
Greece	0.522	0.337	0.185	0.471	0.338	0.133
Iceland	0.393	0.244	0.149	0.348	0.247	0.101
Ireland	0.591	0.331	0.260	0.545	0.339	0.206
Israel	0.501	0.376	0.125	0.469	0.368	0.101
Italy	0.503	0.319	0.184	0.437	0.321	0.116
Japan	0.488	0.336	0.152	0.409	0.332	0.077
Luxembourg	0.464	0.270	0.194	0.420	0.272	0.148
Netherlands	0.424	0.288	0.136	0.391	0.293	0.098
New Zealand	0.454	0.317	0.137	0.396	0.306	0.090
Norway	0.423	0.249	0.174	0.391	0.257	0.134
Poland	0.468	0.305	0.163	0.427	0.308	0.119
Portugal	0.522	0.344	0.178	0.462	0.339	0.123
Russian Federation	0.486	0.428	0.058	0.461	0.420	0.041
Slovak Republic	0.437	0.261	0.176	0.380	0.260	0.120
Slovenia	0.453	0.246	0.207	0.399	0.242	0.157
South Korea	0.342	0.311	0.031	0.316	0.294	0.022
Spain	0.507	0.338	0.169	0.452	0.337	0.115
Sweden	0.441	0.269	0.172	0.385	0.270	0.115
Switzerland	0.372	0.298	0.074	0.328	0.290	0.038
United Kingdom	0.523	0.341	0.182	0.477	0.347	0.130
United States	0.499	0.380	0.119	0.468	0.375	0.093

Source: OECD Inequality Database, accessed October 23, 2013. Data for most OECD countries are for 2010; exceptions include South Korea (from 2011); Ireland, Japan, New Zealand, and Switzerland (from 2009); and the Russian Federation (from 2008).

mid-1980s, but with data going back to the 1970s for a few.[26] Then we discuss trends in the Gini coefficient for pretax and transfer income and the extent to which taxes and transfers lower the Gini coefficient in a broader range of OECD countries. Finally, we compare trends in inequality since the mid-1980s using all three distributional statistics for working-age population and for all ages.

8.4.1.4.1 Trends in Equivalized DHI Gini Coefficients for 10 Rich Nations

Most of the rich nations that have collected comparable, mostly annual, data since the early 1970s have experienced sizeable increases in the Gini coefficient[27] (Figure 8.13). For some countries those increases came in the 1980s (United States, United Kingdom, and the Netherlands), whereas for others they came in the 1990s and early 2000s (Canada, the Nordic countries, and Germany). Inequality trends in these countries can be thought of as following a J or U shape to varying degrees (see Gottschalk and Smeeding, 2000, for further discussion).

In Italy and France, inequality decreased in the 1980s, and since the mid-1990s the Gini coefficient has changed little in either country. Italy's early 1980s declines were, however, offset by increases in the early 1990s (Brandolini and Vecchi, 2011). Most of the rich nations included in Figure 8.6 have experienced relatively small changes in their DHI inequality over the last 10 or 20 years, but many have witnessed marked cyclical fluctuations, particularly the United States, the United Kingdom, and the Nordic countries.

In most cases the rank ordering of countries remains unchanged after nearly 40 years of mostly rising inequality. The most dramatic shifts were undertaken by France, which had the most unequal distribution (among these rich nations) in the mid-1970s and now has a Gini coefficient only modestly higher than that of the Nordic countries. Also, the United Kingdom had among the least unequal distributions in the mid-1970s and has been among the most unequal since the early 1990s. The United States has had the most unequal distribution of income among rich nations since the early 1980s.

Rising inequality in the Nordic countries has produced relative but notable shifts as well. Through the early 1990s, the distribution of income in the Nordic countries was substantially less unequal than it was in other countries; since that time rising inequality in the Nordic countries and stable (France and the Netherlands) or modestly rising

[26] There are 14 OECD countries with S80/S20 and P90/P10 statistics available starting in or before the mid-1980s. For five of those countries there are some data for some years in the 1970s. For several of these countries there is a single year of data for these indicators available for the 1970s.

[27] Data for these countries are collected by the OECD (described in their income distribution database) and by national statistical agencies in some countries. Data for the United Kingdom, Finland, Norway, France, and Germany are from national statistical agencies and are published in Atkinson and Morelli (2012, 2014) (updated by the authors), which provides further detail on the sources. Data for Italy were published in Brandolini and Smeeding (2008, 2009) and since updated by Brandolini (personal communication).

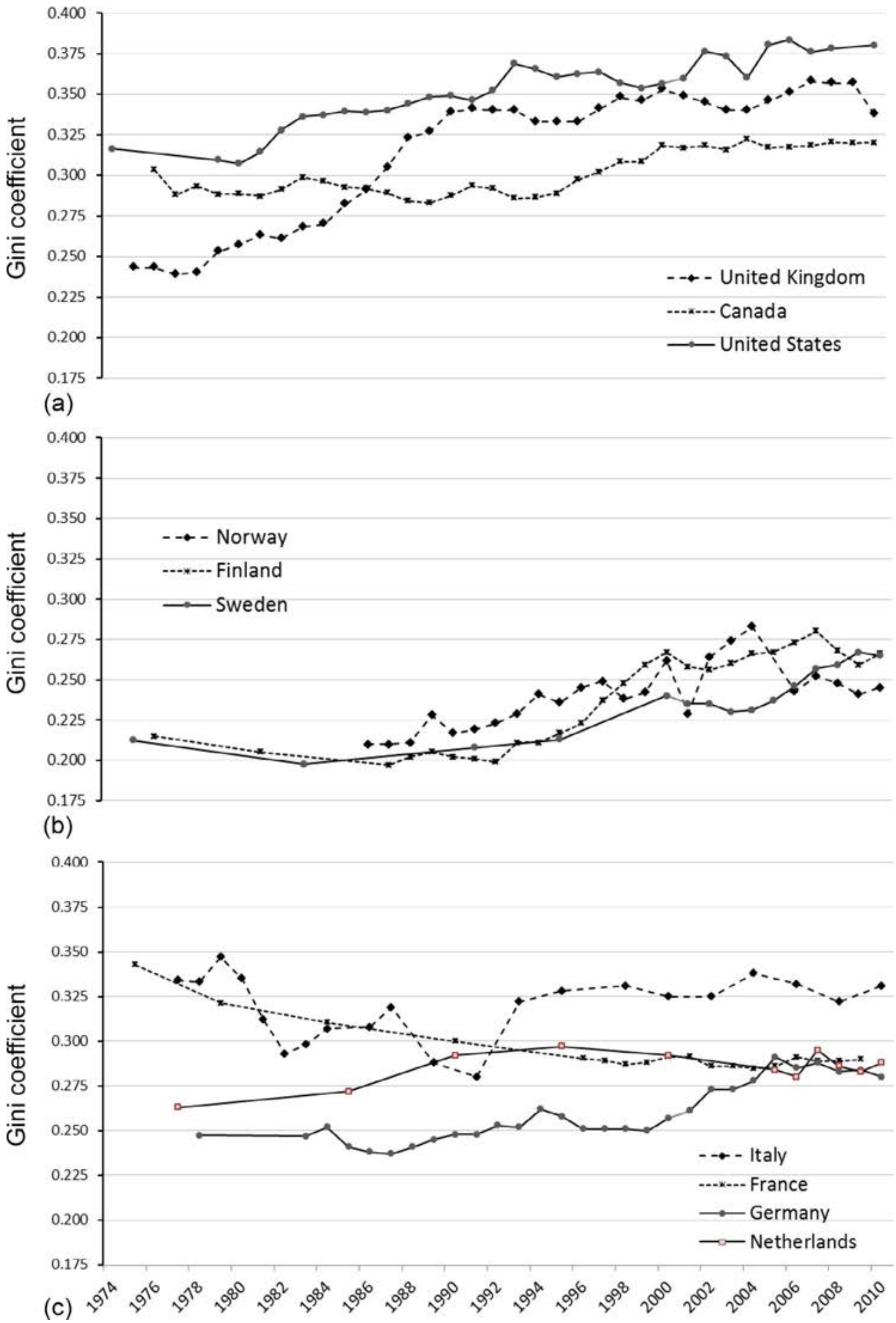


Figure 8.13 Trends in equivalized DHI Gini coefficient in rich countries by country group, OECD, and statistical agency data: Anglo-Saxon countries and the United States (a), Nordic countries (b), and Continental and Southern Europe (c). *Source: OECD income distribution data for Canada, Sweden, and United States.* Inequality Chartbook *(Atkinson and Morelli, 2012, 2014) based on figures published by statistical agencies for remaining countries and updated by authors. Data for Italy from Smeeding and Brandolini (2009), updated by Brandolini.*

inequality (Germany) in other countries has produced some convergence in the inequality levels in continental Europe and the Nordic countries. In Germany, the Gini of DHI rose 14% (from 0.25 to 0.28) over the period.[28] Although the distribution of income was less unequal in the Lander of the former East Germany (EDHI Gini of 0.20 in East Germany and 0.25 in West Germany in 1991), reunification had little impact on the inequality trends for Germany (Fuchs-Schündeln and Schündeln, 2009; Grabka and Kuhn, 2012).

Compared with the early 1980s, the range of inequality measures of these 10 rich countries has become somewhat more compressed. The two nations that previously had the most equal distributions—Sweden and Finland—experienced some of the largest increases in inequality. Around 1980, this set of 10 rich countries had a mean DHI Gini of 0.265 with a variance of 0.0022; around 2010 the mean had increased to 0.30, while the variance had decreased to 0.0017.

8.4.1.4.2 Trends in the S80/S20 and P90/P10 Measures for Equivalized DHI for 14 OECD Countries

A somewhat larger group of countries has been collecting comparable income data at least since the early 1980s (with Denmark, Israel, Japan, Luxembourg, and New Zealand augmenting the 10 rich nations discussed in the previous section).[29] The OECD has analyzed the income surveys from those countries and calculated S80/S20 and P90/P10 ratios. Both of these alternative measures yield largely similar trends in income inequality to what we saw for the Gini coefficient in Figure 8.13.

The share of income received by the top quintile divided by the share of income received by the bottom quintile (S80/S20) has increased in each of these countries since the early 1980s, but some countries experienced larger increases in inequality, and the rank ordering changed somewhat (Figure 8.14). Israel experienced the largest absolute change over this period, with its S80/S20 rising by 2.5, basically matching the United States for top spot with the top fifth of households receiving 7.8 times as much income as those in the bottom fifth. Israel's inequality surge occurred in the late 1990s and early 2000s. Sweden experienced the largest relative increase over the same period—its S80/S20 increased 48%. Canada experienced the smallest increase among these countries: its S80/S20 increased less than 10% higher than its lowest point in the 1980s.

Shifting to an inequality measure that further sharpens the contrast between the top and bottom of income distribution, the P90/P10 interdecile ratio does little to change the trends (Figure 8.15). Similar to the S80/S20 measure, income inequality did increase in each of the countries over this period. For Israel and Japan, the distribution seems to have

[28] For further analysis of trends in German inequality see Grabka and Kuhn (2012), Faik (2012), and Goebel et al. (2010).

[29] France is not included in the OECD series that are available to calculate S80/S20 and P90/P10 ratios.

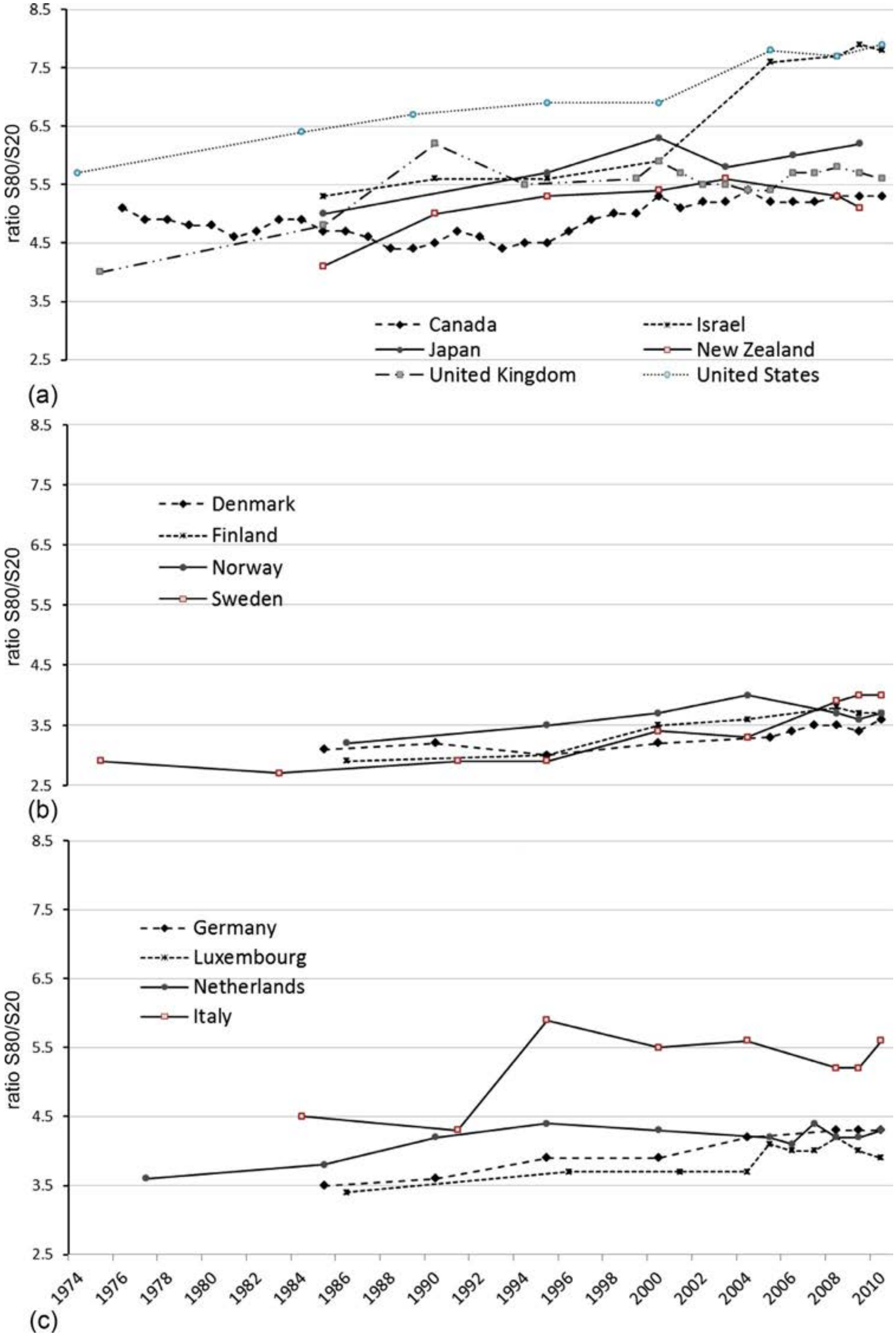


Figure 8.14 Trends in S80/S20 ratio for equivalized DHI by country group based on OECD data: Anglo-Saxon countries, the United States, and others (a); Nordic countries (b); and Continental and Southern Europe (c). *Source: OECD income distribution data.*

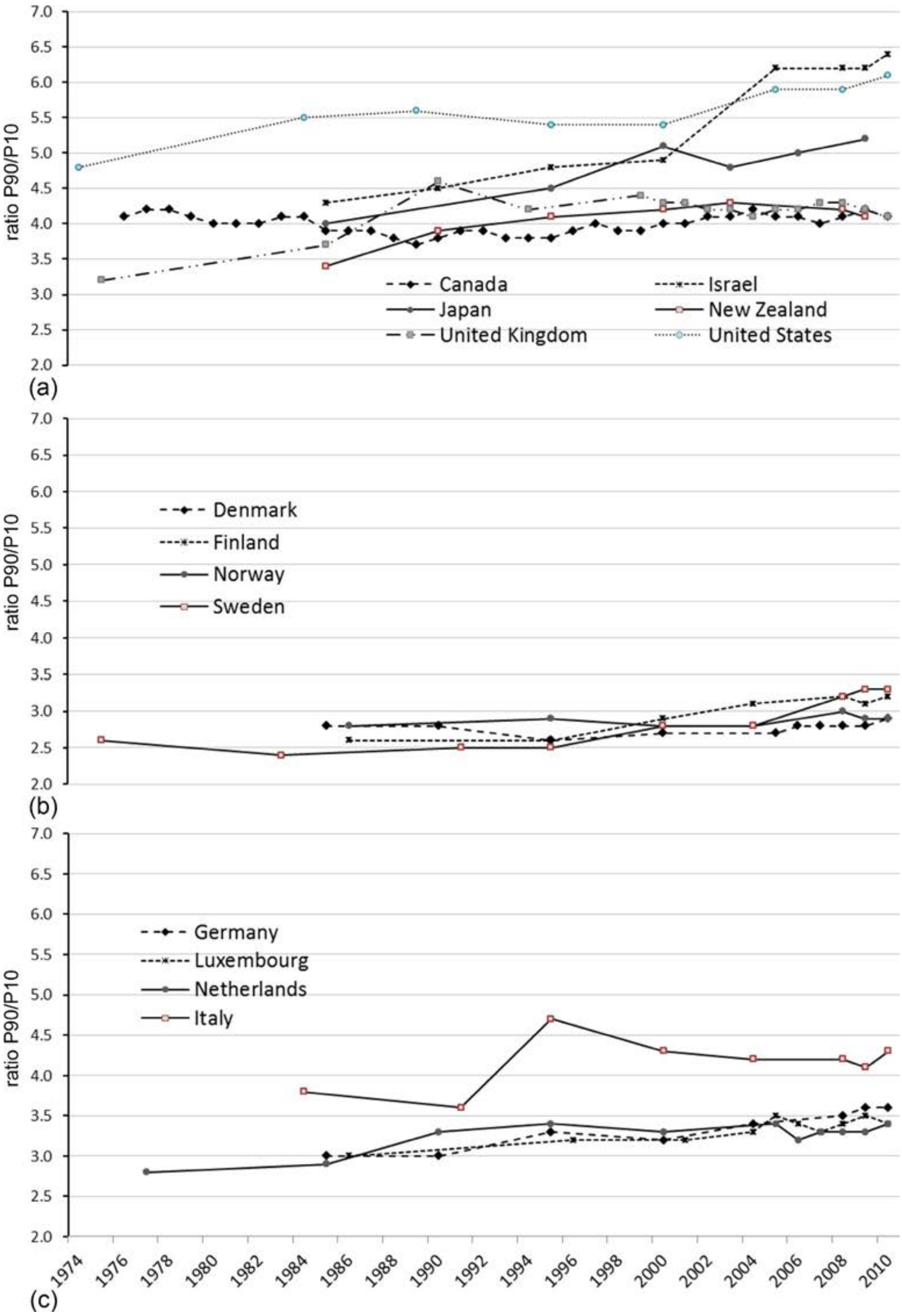


Figure 8.15 Trends in P90/P10 ratios for equivalized DHI by country group based on OECD data: Anglo-Saxon countries, the United States, and others (a); Nordic countries (b); and Continental and Southern Europe (c). *Source: OECD income distribution data.*

grown even more unequal using the P90/10 ratio. By the mid-2000s Israel had supplanted the United States as the most unequal rich nation, with households at the 90th percentile receiving DHIs 6.4 times greater than those at the 10th percentile. In Japan the P90/P10 ratio rose 30% over these three decades.

In most cases, though, the increase in inequality since the early 1980s is equivalent to or somewhat smaller than what is indicated by trends in the S80/S20. In the case of Canada, the 2010 value for the P90/P10 ratio was equal to its 1983 value but 0.4 above its low point in the 1980s. In all of the Nordic and continental European countries (except the Netherlands), the P90/P10 ratio increased less in percentage terms than the S80/S20 did over the same period.

8.4.1.4.3 Trends in Pretax and Transfer Gini Coefficients and the Extent of Redistribution for OECD Countries

Trends in the distribution of pretax and transfer income (using the Gini coefficient) and the extent of redistribution can be explored using the same data, which are available for an expanded set of OECD countries, though for some only since the mid-1990s.

Japan is the country with the largest increase in pretax and transfer inequality among these high-income countries, increasing more than 40% and going from the least unequal distribution in the mid-1980s to one of the most unequal in 2010 (Figure 8.16a). (Figures 8.16 and 8.17 include only countries with data available for the mid-1980s and show the percentage change relative to the mid-1980s base.) Italy also experienced relatively large increases in inequality over this period, with its pretax and transfer Gini increasing 30% (Figure 8.16c). Pretax and transfer inequality increased in most of these countries. In the Anglo-Saxon countries and the United States, the increases were concentrated in the 1980s and early 1990s; in the Nordic and continental European countries Gini coefficients increased most in the early 1990s. The only country that seemed to avoid increased pretax and transfer inequality was the Netherlands. The pretax and transfer Gini actually fell more than 10% in Finland in the late 1990s. For the US, the Netherlands, and Finland, however, data are only available since the mid-1990s. In the Netherlands, increases in the pretax and transfer Gini coefficient in the 1980s were offset by decreases in the late 1990s. New Zealand also experienced declining inequality in the 2000s. A number of countries (including Finland, Israel, and Sweden) have witnessed very little change in pretax inequality over the past 15 years; Gini coefficients have fluctuated only slightly between the mid-1990s and 2010.[30]

Incorporating the influence of taxes and transfers can produce inequality trends that seem quite different, in some cases, than what we see in market or pretax and transfer

[30] For some countries, at least, this finding only holds in income excluding capital gains. In Sweden, for example, inclusion of capital gains income results in sizable increases in pretax and transfer Gini coefficients over this period (Atkinson and Morelli, 2012).

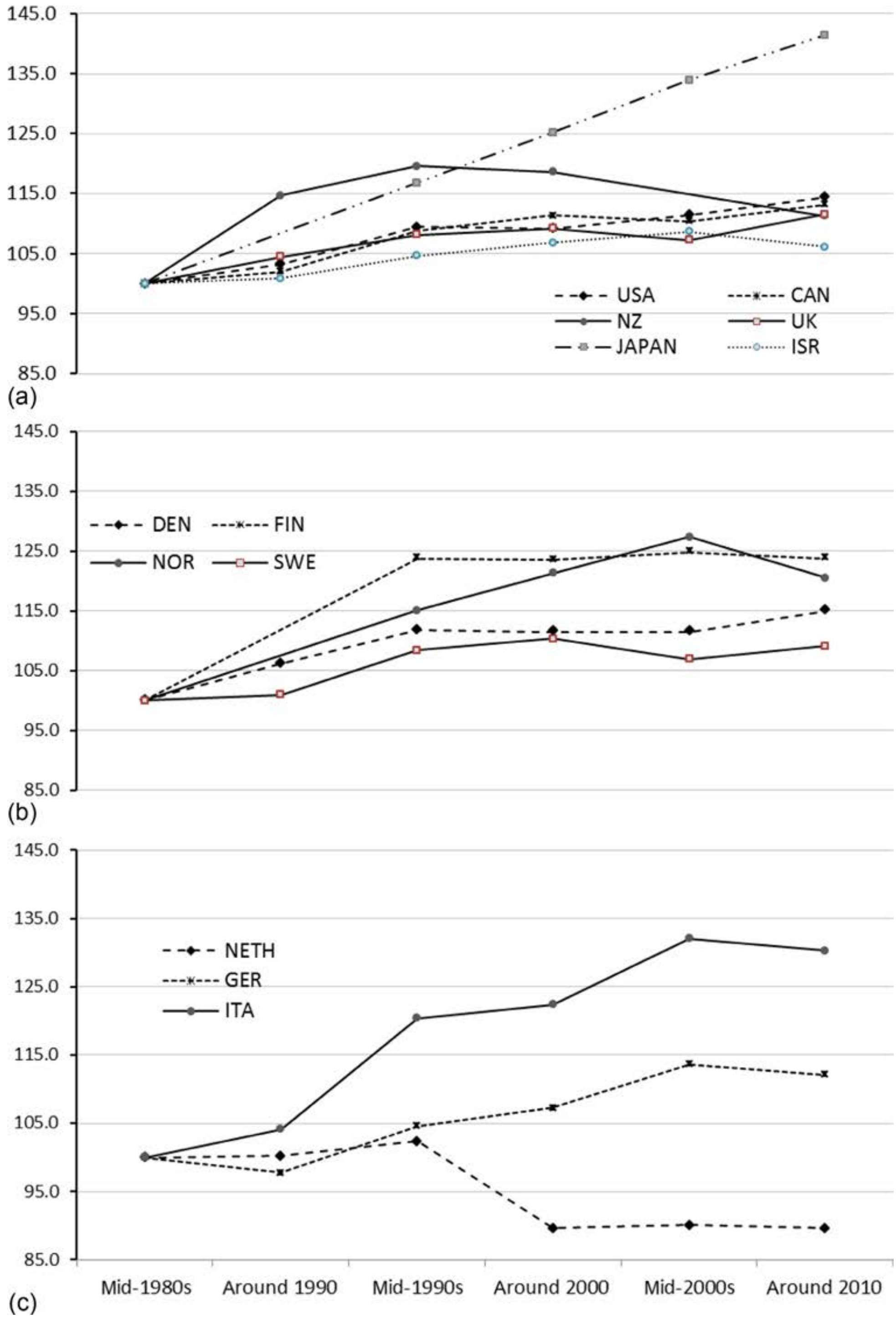


Figure 8.16 Change in pretax and transfer income Gini (mid-1890s = 100) for OECD countries, by country group: Anglo-Saxon countries, the United States, and others (a); Nordic countries (b); and Continental and Southern Europe (c). *Source: OECD Inequality Database, accessed October 23, 2013.*

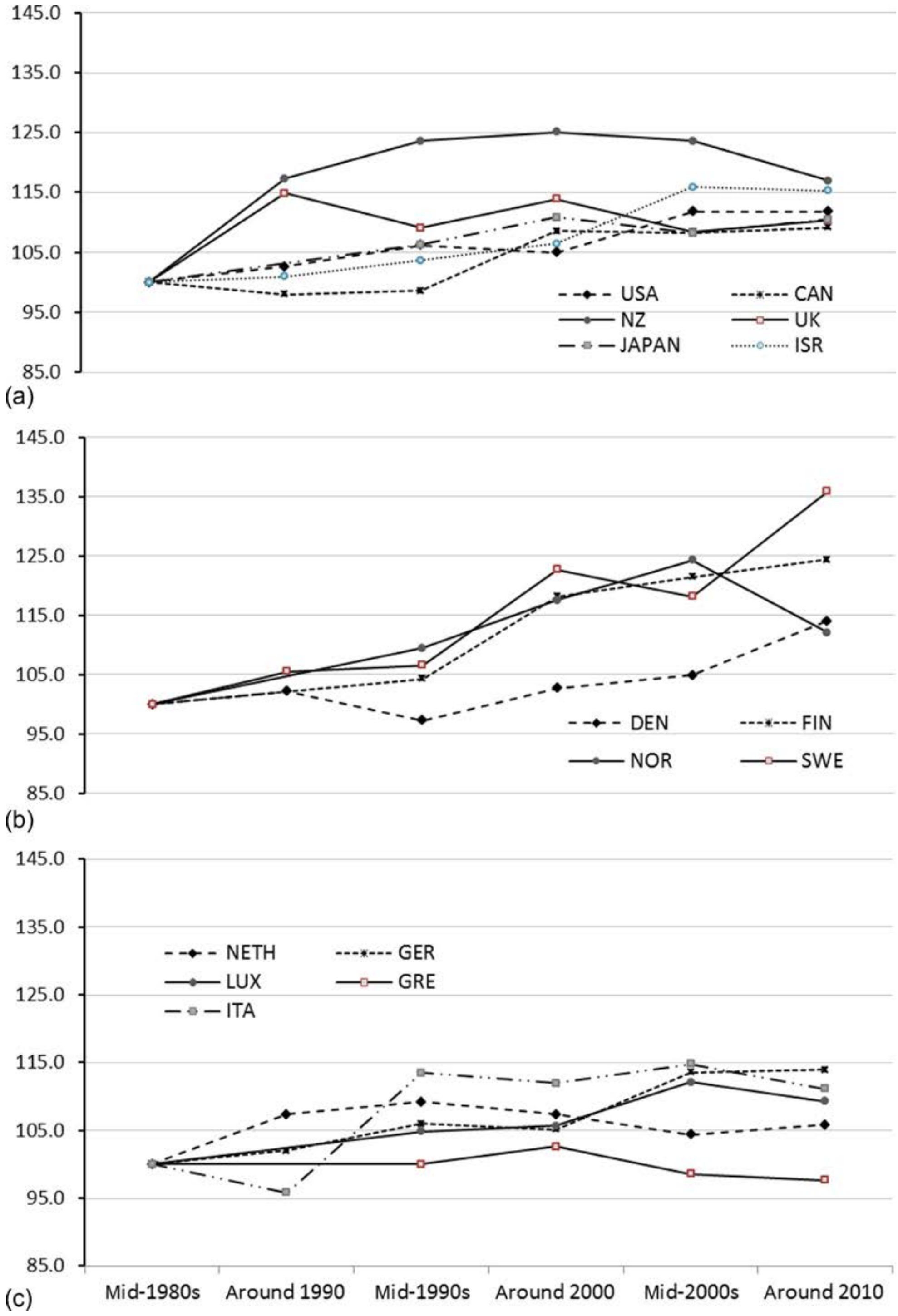


Figure 8.17 Change in disposable household income Gini (mid-1980s = 100) for OECD countries, by country group: Anglo-Saxon countries, the United States, and others (a); Nordic countries (b); and Continental and Southern Europe (c). *Source: OECD Inequality Database, accessed October 23, 2013.*

income. (See Deding and Dall Schmidt, 2002 for an earlier analysis of this issue during the 1990s.) Trends in the Gini coefficients using equalized DHI for the same countries over the same period are shown in Figure 8.17. For some countries trends for the DHI and pretax and transfer Gini coefficients are very similar. The United States, for example, saw the pretax and transfer Gini coefficient increase 14% between the mid-1980s and 2010, while its DHI Gini coefficient increased 12%. The United States is one of the countries for which excluding trends for the 1970s substantially understates its increase in inequality; between the mid-1970s and 2010 the U.S. pretax and transfer Gini increased 23% and its DHI Gini increased 20%.

Denmark, Finland (Figure 8.17b), and Germany (Figure 8.17c) also saw similar increases in the Gini coefficient before and after the inclusion of taxes and transfers. For a number of countries, though, the inclusion of taxes and transfers produces markedly different trends in inequality. For Canada, Japan, Italy, Norway, and the United Kingdom, rising inequality in the distribution of income is blunted once taxes and transfers are included. Japan and Italy, the countries with the largest increases in pretax and transfer inequality in Figure 8.16, experienced increases in their DHI inequality only one-quarter and one-third as large, respectively. The opposite is the case for Sweden, the Netherlands, Israel, and New Zealand, which experienced larger increases in inequality after including taxes and transfers. In the case of Sweden, the Gini for pretax and transfer income increased 9% (from 0.40 to 0.44) between the mid-1980s and 2010, while the Gini for DHI increased 36% (from 0.19 to 0.27).

The differences in the trends illustrated in Figures 8.16 and 8.17 are partly a result of the evolution of the tax and transfer systems in these countries. (As mentioned previously, changes in the age of the population and other demographic and policy factors can influence these trends as well.) Figure 8.18 shows how the extent to which tax and transfers reduce the Gini coefficients for pretax and transfer incomes has changed over this period. The most striking pattern in Figure 8.18 is the dramatic and sustained increase in tax and transfer "redistribution" in most rich nations from the mid-1970s through the mid-1990s, which was followed by steady declines in the decade and a half since. The Anglo-Saxon (Figure 8.18a) and Nordic countries (Figure 8.18b) in particular followed an inverse U-shaped pattern, with redistributive efforts increasing between the mid-1970s and mid-1990s but declining after that point. Since around 2000, taxes and transfers also have played a smaller role in reducing market income inequality in Israel.

In some countries the redistributive impact did not subside in the 1990s. Japan (Figure 8.18a) and Italy (Figure 8.18c) both experienced steady increases in redistribution from the 1990s through the late 2000s. The impact of redistribution has fluctuated less in the United States than in most other high-income countries. Increased redistribution in Canada and Japan, though, has shifted the United States from having one of the lowest levels of redistribution to having the lowest among rich nations. (See Caminada et al., 2012; Immervoll and Richardson, 2011; Wang and Caminada, 2011, for more detailed

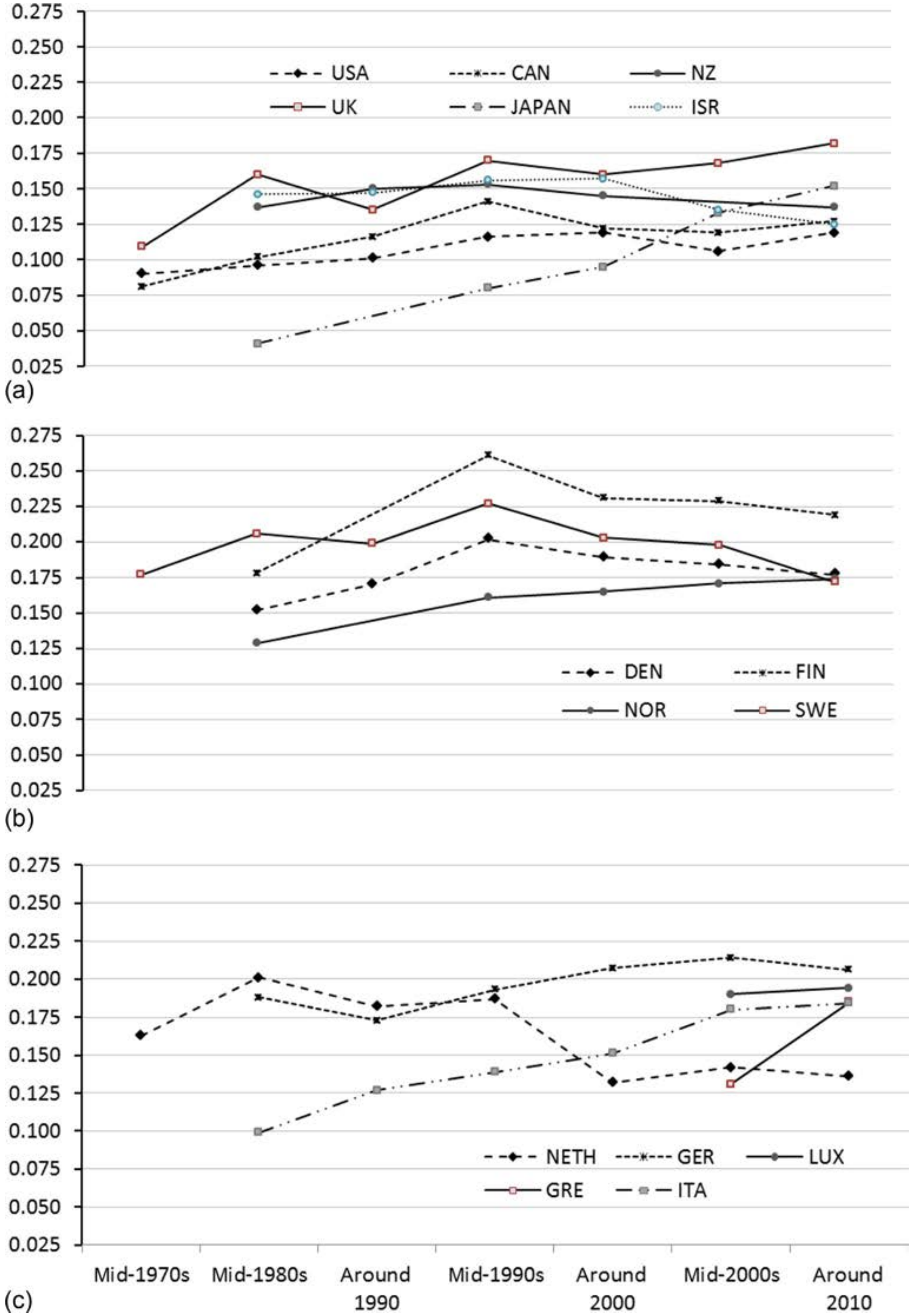


Figure 8.18 Reduction in Gini coefficient due to taxes and transfers trends for OECD countries, by country group: Anglo-Saxon countries, the United States, and others (a); Nordic countries (b); and Continental, Southern, and Eastern Europe (c). *Source: OECD Inequality Database, accessed October 23, 2013.*

discussion of the specific policies and their contribution to reducing market inequality in OECD and LIS countries.) Gini coefficients for pretax and transfer income, DHI, and the difference between the two for the OECD countries are shown in Table 8.5.

8.4.1.4.4 Comparing Trends in DHI Inequality for All Ages and the Working-Age Population

We previously described how age composition is important in understanding how taxes and transfers affect cross-national rankings of income inequality. We can use the same OECD data to evaluate trends in income distribution statistics for the working-age population and contrast them with trends for the overall population. Table 8.6 includes S80/S20 and P90/P10 ratios, as well as Gini coefficients using equivalized DHI for a selection of years between the mid-1980s and 2010 for high-income OECD countries.

Over the entire 25-year period, the distribution of income grew even more unequal among the working-age population in almost every country. The largest differences can be seen among the Nordic countries. In Norway and Sweden, the P90/P10 ratio increased 20% and 15% more among the working-age than for the overall population, respectively, between the mid-1980s and 2010 (panel A). Smaller differences can be seen for the United States, United Kingdom, and Canada, which saw inequality increase between 4% and 8% more among the working-age than among all ages combined. Israel is the only country to see larger increases among the overall population than the working-age, although New Zealand also saw larger increases among the overall population after the mid-1990s. For some of the countries, though, there is no notable difference in the inequality trends between the working-age and the overall population at any point, or at least in more recent years.

The S80/S20 ratio measure yields a strikingly similar pattern of results (panel B) as the P90/P10 ratio, but differences in Gini coefficient trends between the age groups (panel C) are more muted. Norway and Denmark saw DHI Gini coefficients increase 10% and 5% more, respectively, among the working-age than the overall population between 1985 and 2010. In most countries, however, trends in the Gini coefficient were only modestly greater among the working-age population. The tails of the distribution have a greater impact on the S80/S20 and P90/P10 measures than they do on the Gini coefficient and seem to be particularly relevant to understanding any differences in inequality trends for different age groups.

8.4.2 Top Incomes

8.4.2.1 Introduction

The first empirical section of this chapter focused on incomes at the bottom of the distribution relative to the poverty line. The previous section discussed trends in the overall distribution of income (e.g., Gini coefficients), suggesting that the distribution of income has become more unequal in most countries since the 1970s. This section shifts attention to the top of the distribution. Top incomes deserve a separate discussion because top

Table 8.5 Gini index for market income and post-tax/transfer income and the extent of redistribution

	Mid-1970s	Mid-1980s	Around 1990	Mid-1990s	Around 2000	Mid-2000s	Around 2010
Panel 1: Market income							
Australia				0.467	0.476	0.465	0.469
Austria						0.464	0.479
Belgium						0.482	0.478
Canada	0.385	0.395	0.403	0.43	0.44	0.436	0.447
Czech Republic				0.442	0.472	0.461	0.449
Denmark		0.373	0.396	0.417	0.416	0.416	0.429
Estonia						0.485	0.487
Finland		0.387		0.479	0.478	0.483	0.479
France				0.473	0.49	0.485	0.505
Germany		0.439	0.429	0.459	0.471	0.499	0.492
Greece						0.471	0.522
Hungary							
Iceland						0.373	0.393
Ireland						0.504	0.591
Israel		0.472	0.476	0.494	0.504	0.513	0.501
Italy		0.386	0.402	0.465	0.472	0.51	0.503
Japan		0.345		0.403	0.432	0.462	0.488
Luxembourg						0.467	0.464
Netherlands	0.426	0.473	0.474	0.484	0.424	0.426	0.424
New Zealand		0.408	0.468	0.488	0.484		0.454
Norway		0.351		0.404	0.426	0.447	0.423
Poland						0.521	0.468
Portugal						0.498	0.522
Slovak Republic						0.462	0.437
Slovenia						0.448	0.453
South Korea						0.33	0.342
Spain						0.463	0.507
Sweden	0.389	0.404	0.408	0.438	0.446	0.432	0.441
Switzerland							0.372
Turkey							
United Kingdom	0.378	0.469	0.49	0.507	0.512	0.503	0.523
United States	0.406	0.436	0.45	0.477	0.476	0.486	0.499
Russian Federation							0.486
Panel 2: Post-tax and transfer (DHI)							
Australia				0.309	0.317	0.315	0.334
Austria						0.26	0.267
Belgium						0.269	0.262
Canada	0.304	0.293	0.287	0.289	0.318	0.317	0.32

Table 8.5 Gini index for market income and post-tax/transfer income and the extent of redistribution—cont'd

	Mid-1970s	Mid-1980s	Around 1990	Mid-1990s	Around 2000	Mid-2000s	Around 2010
Czech Republic			0.232	0.257	0.26	0.259	0.256
Denmark		0.221	0.226	0.215	0.227	0.232	0.252
Estonia						0.337	0.319
Finland		0.209		0.218	0.247	0.254	0.26
France				0.277	0.287	0.288	0.303
Germany		0.251	0.256	0.266	0.264	0.285	0.286
Greece	0.424	0.345		0.345	0.354	0.34	0.337
Hungary			0.273	0.294	0.293	0.291	0.272
Iceland						0.269	0.244
Ireland						0.315	0.331
Israel		0.326	0.329	0.338	0.347	0.378	0.376
Italy		0.287	0.275	0.326	0.321	0.33	0.319
Japan		0.304		0.323	0.337	0.329	0.336
Luxembourg		0.247		0.259	0.261	0.277	0.27
Netherlands	0.263	0.272	0.292	0.297	0.292	0.284	0.288
New Zealand		0.271	0.318	0.335	0.339	0.335	0.317
Norway		0.222		0.243	0.261	0.276	0.249
Poland						0.326	0.305
Portugal						0.373	0.344
Slovak Republic						0.275	0.261
Slovenia						0.245	0.246
South Korea						0.306	0.311
Spain						0.324	0.338
Sweden	0.212	0.198	0.209	0.211	0.243	0.234	0.269
Switzerland							0.298
Turkey		0.434		0.49		0.43	0.411
United Kingdom	0.269	0.309	0.355	0.337	0.352	0.335	0.341
United States	0.316	0.34	0.349	0.361	0.357	0.38	0.38
Russian Federation							0.428
Panel 3: Redistribution (market Gini less DHI Gini)							
Australia				0.158	0.159	0.15	0.135
Austria						0.204	0.212
Belgium						0.213	0.216
Canada	0.081	0.102	0.116	0.141	0.122	0.119	0.127
Czech Republic				0.185	0.212	0.202	0.193
Denmark		0.152	0.17	0.202	0.189	0.184	0.177
Estonia						0.148	0.168
Finland		0.178		0.261	0.231	0.229	0.219

Continued

Table 8.5 Gini index for market income and post-tax/transfer income and the extent of redistribution—cont'd

	Mid-1970s	Mid-1980s	Around 1990	Mid-1990s	Around 2000	Mid-2000s	Around 2010
France				0.196	0.203	0.197	0.202
Germany		0.188	0.173	0.193	0.207	0.214	0.206
Greece						0.131	0.185
Hungary							
Iceland						0.104	0.149
Ireland						0.189	0.26
Israel		0.146	0.147	0.156	0.157	0.135	0.125
Italy		0.099	0.127	0.139	0.151	0.18	0.184
Japan		0.041		0.08	0.095	0.133	0.152
Luxembourg						0.19	0.194
Netherlands	0.163	0.201	0.182	0.187	0.132	0.142	0.136
New Zealand		0.137	0.15	0.153	0.145		0.137
Norway		0.129		0.161	0.165	0.171	0.174
Poland						0.195	0.163
Portugal						0.125	0.178
Slovak Republic						0.187	0.176
Slovenia						0.203	0.207
South Korea						0.024	0.031
Spain						0.139	0.169
Sweden	0.177	0.206	0.199	0.227	0.203	0.198	0.172
Switzerland							0.074
Turkey							
United Kingdom	0.109	0.16	0.135	0.17	0.16	0.168	0.182
United States	0.09	0.096	0.101	0.116	0.119	0.106	0.119
Russian Federation							0.058

Source: OECD Inequality Database, accessed October 23, 2013.
Note: For most OECD countries, "Around 2010" is for the year 2010, with some exceptions: for South Korea, data are for 2011; for Hungary, Ireland, Japan, New Zealand, Switzerland, and Turkey, data are for 2009; and for the Russian Federation, data are for 2008.

income measures taken from household surveys are typically less accurate because of both sampling and nonsampling errors.

The main objective of this section is to discuss the trends of the so-called top income shares as computed from administrative tax statistics. Different from Chapter 7, we focus here on the investigation of the four decades since 1970. Moreover, we mainly describe here the trends in top shares, leaving out a discussion of what may have driven such trends (see Part III of this volume). The methodological issues affecting the comparison of trends over time and across countries also define a substantial part of this section. Indeed, we start

Table 8.6 Comparing trends in equivalized DHI inequality for all ages and for the working-age population, by measure by country

	All ages				Working-age population				Change (%) Mid-1980s to 2010		1995–2010		2005–2010	
	1985	1995	2005	2010	1985	1995	2005	2010	All ages	Working age	All ages	working age	All ages	working age
Panel A: P90/P10 ratio														
Canada	3.9	3.8	4.1	4.1	4	4	4.4	4.4	5%	10%	8%	10%	0%	0%
Denmark	2.8	2.6	2.7	2.9	2.5	2.5	2.7	2.9	4%	16%	12%	16%	7%	7%
Finland	2.6	2.6	3.1	3.2	2.6	2.7	3.1	3.3	23%	27%	23%	22%	3%	6%
Germany	3	3.3	3.4	3.6	2.9	3.2	3.5	3.6	20%	24%	9%	13%	6%	3%
Israel	4.3	4.8	6.2	6.4	4.4	4.8	6.1	6.1	49%	39%	33%	27%	3%	0%
Italy	3.8	4.7	4.2	4.3	3.6	4.6	4	4.5	13%	25%	−9%	−2%	2%	13%
Japan	4	4.5	5	5.2	4	4.5	4.9	5.3	30%	33%	16%	18%	4%	8%
Luxembourg	3	3.2	3.5	3.4	2.9	3.2	3.6	3.4	13%	17%	6%	6%	−3%	−6%
Netherlands	2.9	3.4	3.4	3.4	3	3.5	3.5	3.6	17%	20%	0%	3%	0%	3%
New Zealand	3.4	4.1	4.3	4.1	3.4	4.3	4.6	4.1	21%	21%	0%	−5%	−5%	−11%
Norway	2.8	2.9	2.8	2.9	2.6	2.8	2.9	3.2	4%	23%	0%	14%	4%	10%
Sweden	2.4	2.5	2.8	3.3	2.3	2.6	2.9	3.5	38%	52%	32%	35%	18%	21%
United Kingdom	3.7	4.2	4.2	4.1	3.7	4.1	4.4	4.4	11%	19%	−2%	7%	−2%	0%
United States	5.5	5.4	5.9	6.1	5.3	5.3	5.7	6.1	11%	15%	13%	15%	3%	7%
Panel B: S80/S20 ratio														
Canada	4.7	4.5	5.2	5.3	4.8	4.8	5.6	5.6	13%	17%	18%	17%	2%	0%
Denmark	3.1	3	3.3	3.6	3	2.9	3.3	3.7	16%	23%	20%	28%	9%	12%
Finland	2.9	3	3.6	3.7	2.9	3.1	3.7	3.9	28%	34%	23%	26%	3%	5%
Germany	3.5	3.9	4.2	4.3	3.5	3.9	4.4	4.4	23%	26%	10%	13%	2%	0%
Greece	6.2	6.2	5.8	6	6.1	5.8	5.9	6.1	−3%	0%	−3%	5%	3%	3%
Israel	5.3	5.6	7.6	7.8	5.4	5.6	7.5	7.6	47%	41%	39%	36%	3%	1%
Italy	4.5	5.9	5.6	5.6	4.4	5.9	5.5	5.8	24%	32%	−5%	−2%	0%	5%
Japan	5	5.7	6	6.2	5	5.5	5.9	6.2	24%	24%	9%	13%	3%	5%
Luxembourg	3.4	3.7	4.1	3.9	3.4	3.7	4.3	4	15%	18%	5%	8%	−5%	−7%
Netherlands	3.8	4.4	4.2	4.3	3.9	4.5	4.3	4.6	13%	18%	−2%	2%	2%	7%
New Zealand	4.1	5.3	5.6	5.1	4	5.4	5.9	5	24%	25%	−4%	−7%	−9%	−15%
Norway	3.2	3.5	4	3.7	3	3.5	4.2	4	16%	33%	6%	14%	−8%	−5%

Continued

Table 8.6 Comparing trends in equivalized DHI inequality for all ages and for the working-age population, by measure by country—cont'd

	All ages				Working-age population				Change (%) Mid-1980s to 2010		1995–2010		2005–2010	
	1985	1995	2005	2010	1985	1995	2005	2010	All ages	Working age	All ages	working age	All ages	working age
Sweden	2.7	2.9	3.3	4	2.6	3.1	3.4	4.3	48%	65%	38%	39%	21%	26%
United Kingdom	4.8	5.5	5.4	5.6	5	5.8	5.8	6.2	17%	24%	2%	7%	4%	7%
United States	6.4	6.9	7.8	7.9	6.1	6.7	7.6	7.9	23%	30%	14%	18%	1%	4%
Panel C: Gini coefficient														
Canada	0.293	0.289	0.317	0.32	0.291	0.293	0.322	0.324	9%	11%	11%	11%	1%	1%
Denmark	0.221	0.215	0.232	0.252	0.209	0.206	0.227	0.248	14%	19%	17%	20%	9%	9%
Finland	0.209	0.218	0.254	0.26	0.206	0.224	0.253	0.263	24%	28%	19%	17%	2%	4%
Germany	0.251	0.266	0.285	0.286	0.246	0.267	0.288	0.285	14%	16%	8%	7%	0%	−1%
Greece	0.345	0.345	0.34	0.337	0.344	0.336	0.337	0.338	−2%	−2%	−2%	1%	−1%	0%
Israel	0.326	0.338	0.378	0.376	0.317	0.329	0.374	0.368	15%	16%	11%	12%	−1%	−2%
Italy	0.287	0.3258	0.3295	0.319	0.284	0.3235	0.3243	0.321	11%	13%	−2%	−1%	−3%	−1%
Japan	0.304	0.323	0.329	0.336	0.304	0.319	0.323	0.332	11%	9%	4%	4%	2%	3%
Luxembourg	0.247	0.259	0.277	0.27	0.239	0.261	0.281	0.272	9%	14%	4%	4%	−3%	−3%
Netherlands	0.272	0.297	0.284	0.288	0.273	0.298	0.285	0.293	6%	7%	−3%	−2%	1%	3%
New Zealand	0.271	0.335	0.335	0.317	0.264	0.329	0.329	0.306	17%	16%	−5%	−7%	−5%	−7%
Norway	0.222	0.243	0.276	0.249	0.211	0.237	0.284	0.257	12%	22%	2%	8%	−10%	−10%
Sweden	0.198	0.211	0.234	0.269	0.195	0.216	0.236	0.27	36%	38%	27%	25%	15%	14%
United Kingdom	0.309	0.337	0.335	0.341	0.305	0.334	0.335	0.347	10%	14%	1%	4%	2%	4%
United States	0.34	0.361	0.38	0.38	0.329	0.351	0.373	0.375	12%	14%	5%	7%	0%	1%

Source: OECD Income Distribution Database, accessed November 8, 2013. Some data for some countries are from the following years: Finland, 1986, 2004; Greece, 1986, 1994; Italy, 1984, 2004; Japan, 2006, 2009; Luxembourg, 1986, 1996; New Zealand, 2003, 2009; Norway, 1986, 2004; Sweden, 1983, 2004; United Kingdom, 1994; United States, 1984.

here with an overview of the main features and limitations of the data with the objective of highlighting how the latter can affect the comparability of the top shares over time and across countries. Where possible, we illustrate how income at the top can be decomposed by different sources, highlighting the role of capital and wage incomes. Similarly, we provide a brief description of the impact of fiscal policy on the top income shares after taxes. Differences in tax systems can affect differently the level as well the trend of top shares across countries. Finally, we discuss how we can complement the two sources of information (tax and survey statistics) to improve our understanding of the evolution of income inequality.

The analysis in this section uses data on total income of the families, tax units, and individuals above the 99th percentile of the distribution. Therefore, unlike Chapter 7, this chapter does not focus on different income groups within the top decile. The data are collected and assembled from tax statistics, available from the World Top Incomes Database (WTID) by Alvaredo et al. (2012). The database is the result of years of work in a line of research initiated by Frankel and Herzfeld (1943)[31] and Kuznets (1953), revived by Piketty (2001), and carried on in subsequent collective works directed by Atkinson and Piketty (2007, 2010), who pulled together a number of contributions from different authors.[32]

Motivations for the surge of interest in incomes at the top of the distribution vary. On one hand, the WTID database constitutes a unique source of information covering most of the twentieth century (and in a few cases the beginning of the twenty-first century as well). As shown in great detail in Chapter 7, this is a crucial advantage for studies of income distribution, which are usually plagued by data limitations.

On the other hand, the analysis of top income shares helps to offer a better understanding of the post-1970 dynamics of income distribution and its determinants. First, the share of total income captured by a tiny minority of the population in many advanced countries has been increasing continuously since the 1980s, and this has fueled concerns about the social inclusiveness of economic growth. In the United States, "the top 1 percent captured 58 percent of real economic growth per family" during the 1976–2007 period (Atkinson et al., 2011, p. 8). Findings such as these likely motivated the managing director of the IMF, Christine Lagarde, to refer to inequality and the inclusiveness of growth as one of the three future challenges of the global economy that the IMF aims to address.[33]

[31] Although Kuznets is often considered to be the pioneer of this stream of literature, Alvaredo and Atkinson (2010) noted that "Frankel and Herzfeld (1943) published estimates of the European income distribution in South Africa based on the income tax returns, but making use of control totals from the census of population and from the national accounts. Their use of external information to complement income tax data pre-dated by ten years the study of upper income groups in the United States by Kuznets (1953)."

[32] The reader is directed to the WTID Web site at http://topincomes.g-mond.parisschoolofeconomics.eu/ for a complete list of the sources of data.

[33] "A better financial system" and overcoming economic and financial crisis are the two other points. This refers to the Annual Meetings Speech in Tokyo on October 12, 2012 (http://www.imf.org/external/np/speeches/2012/101212a.htm).

Second, understanding the dynamics of the share of total income of the upper-income brackets may be crucial to understanding changes in the overall income distribution. This has been shown empirically by Leigh (2007) and Smeeding and Thompson (2011) and discussed more formally by Atkinson (2007) and Alvaredo (2011). As in Chapter 7, we recognize here that the relationship between top shares and other income inequality measures may well be changing over time. In particular, we exploit disaggregated evidence for different decades to show that such a relationship has weakened since the 1990s. This differentiates our conclusions from those of Chapter 7, calling for extra prudence in using top shares as a proxy for the overall income distribution as obtained from household surveys. Furthermore, academic research has shown that the information contained within standard surveys hardly captures incomes above the 99th percentile so that top income shares can be potentially used to adjust available measures of overall inequality such as the Gini coefficient, discussed in the previous section.

Third, top income shares have been particularly useful to studies of important issues in public economics, such as the elasticity of reported income to tax changes, the extent of income shifting and tax avoidance, and, more generally, behavioral responses to changes in taxation. Finally, the new empirical evidence on top shares gave the economic profession a new challenge: conventional explanations of rising income inequality since the end of Bretton Woods system, such as the skill-biased technological change and globalization forces, are no longer sufficient to explain the evolution of top income shares across different developed countries.

8.4.2.2 Data and Methodology

As mentioned above, our analysis makes use of the WTID for 21 countries since 1970.[34] In general, the series are constructed using tax statistics, and they make use of gross types of income (e.g., in the United States, the gross market income is defined before deductions, individual income taxes, payroll taxes, and all kinds of government transfers).

Top income shares are mostly calculated from detailed, historically tabulated income tax statistics. Alternatively, tax administration microdata are also increasingly used, especially for the last decades of the twentieth century. Information contained within the tax statistics then is combined with control totals for population and income. Essentially, tax statistics provide the total income and the total number of tax units for given income ranges and allow us to compare these values with the totals in the economy.[35] It is important to note that when using group tabulations data, the precise share of income accruing

[34] The countries are Australia, Canada, China, Denmark, Finland, France, Germany, India, Ireland, Italy, Japan, the Netherlands, New Zealand, Norway, Portugal, South Africa, Spain, Sweden, Switzerland, the United Kingdom, and the United States.

[35] This refers exclusively to those who file taxes, which can be quite a small portion of the population, especially in earlier years of the twentieth century. The assumption is that the top of the distribution is always sufficiently covered.

to a specific percentile within the top decile is obtained through interpolation techniques because the ranges of tax units within tabulations do not necessarily coincide with the percentage of the population for which we would like to assemble data. Interpolation is commonly applied using distributional assumptions about the top tail of income distribution (e.g., Pareto distribution) or, alternatively, by computing lower and upper bounds for every share (e.g., the actual share can be obtained using the mean-split histogram, as used by Atkinson, 2005).

Broadly speaking, the choice between these two different interpolation techniques does not affect the substance of the results, and interpolation errors have been generally proven to be negligible. This is particularly true when the information within the grouped tabulations is detailed and of high quality.[36] While choice of interpolation technique does not seem to be crucial, other factors may substantially influence the accuracy of estimates of top income shares and the comparability of levels and trends across countries.[37]

8.4.2.2.1 Caveats and Limitations to the Data

Although top shares are calculated with similar methodologies across countries, there are a number of caveats that are important to consider.[38] This section summarizes and extends the discussion of the methodology for the derivation of top shares data found in previous publications. Differences and changes in methodology may (or may not) affect the comparability of data across countries as well as over time, even within individual country-specific series. Understanding the relevance of these issues is the focus of the following subsections.

Reliance on tax statistics raises a number of important questions and concerns about the construction of top income shares. First, the income definition is tailored to follow administrative requirements, implying that the definitions of income, income unit, and so on, do not necessarily coincide with the preferred definitions used for research purposes.[39] Administrative criteria may also differ across countries or change over time (e.g., changes to tax legislation such as income sources subjected to taxation, tax units), generating comparability issues.

Calculating comparable top share series also requires consistency between the numerator (total top income) and the denominator (total income in the economy). However, the control total for income is calculated in different ways across countries and over time, in turn affecting the comparability of data. In addition, economic agents have incentives

[36] See Atkinson (2005) for a detailed discussion.

[37] Atkinson et al. cover these issues extensively in their 2011 paper in the *Journal of Economic Literature*.

[38] The limitations of income tax data have been discussed extensively by Leigh (2009), Atkinson et al. (2009, 2011), and Burkhauser et al. (2012b).

[39] Ideally, the Haig-Simons definition of income is preferred. This includes accruing capital gains and losses (not only realized), imputed rents, and fringe employment benefits.

to change their behavior to minimize their tax liabilities and most probably understate their true income (e.g., tax avoidance, tax evasion, income shifting), and these incentives may vary with income and across tax systems.

Finally, the series are largely concerned with gross income before taxes so that the effective change in income inequality after taxes is dependent on changes in the effective tax rate of top income brackets. This has certainly changed dramatically over time and might not have followed a similar pattern in all countries. This is an issue of crucial importance, although it is much less debated because of data limitations. This section addresses these issues separately.

Despite these concerns, it is worth noting that the literature on top income shares also has highlighted the potential and the strength of these data, generally concluding that these problems can be attenuated. The country-specific series are usually obtained from the same sources over time, and we can easily identify breaks that may affect the measurement as well as indicate the direction and magnitude of the potential change. In addition, our focus on the post-1970 period allows us to have both better data and better documentation to deal with these issues to a satisfactory level. For an analysis on the very long run (since approximately 1750), we direct the reader to Chapter 7.

8.4.2.2.2 Definition of the Control Total for Income

Every top share is a fraction between the incomes accruing to a specific top income group with respect to the total income in the economy. The two definitions of income have to be consistent, and there are different ways to come up with an estimate of the total pretax income in the economy.

As detailed by Atkinson et al. (2011), and illustrated within the previous chapter as well, one possible approach is to subtract specific categories of income from the total personal income within the national accounts. This is done to come as close as possible to the income definition reported in the tax statistics (occasionally proportionally adjusted). This is the original approach pioneered by Frankel and Herzfeld (1943), then by Kuznets in 1953, and later adopted by Piketty in 2001 and used by most of the countries in the WTID.[40] The alternative approach is to inflate the total income that is reported in the tax statistics to correct the missing income of individuals who do not file a tax return (as done for the series for the United Kingdom, Finland, the Netherlands, Sweden after 1942, Switzerland after 1971, and the United States after 1944). Whereas the first approach makes use of external control of income from national accounts sources, the second approach deals with sources of income that are mainly internal to the tax statistics.[41]

[40] Australia, Canada, Germany, India, Ireland, Italy, Japan, New Zealand, Norway, South Africa, and Sweden before 1942; Switzerland before 1971; and the United States before 1944.

[41] It is worth noting that additional information (external from tax sources) can also be used to estimate the income of nonfilers (this is, for example, the case of the United Kingdom, as described by Atkinson, 2005).

In a few cases where national accounts are not available (this is especially true for earlier years), total income is estimated from the full population households survey (this was done for China) or as a share of gross domestic product (this is the case for initial observations of Spanish and Portuguese top incomes).

As should be expected, these methodological differences may affect the level of the series as well as the trend. In particular, the comparison of cross-country trends can be affected when different methodologies are systematically applied by different countries. For the case of the United Kingdom, Atkinson (2007) documents how the ratio between total income based on tax statistics and total income from national accounts[42] has declined over time, falling from 0.9 at the beginning of the century to 0.85 in the last years of the century. Assuming that the ratio decreased at a constant rate, we could obtain a rough estimate of the dynamics of the top income shares based on the control totals using national accounts. Despite the documented minor change, the impact on the trend of top shares can be seen over time. The gap between the top shares reported in the WTID (using control total estimated from the tax statistics) and the ones we estimate based on a different control total rises from 1 percentage point in 1970 to 2 percentage points in 2000. In the case of the United Kingdom, the two different approaches yield very similar trends over time, but the magnitude of the increase—whether the top 1% share rose 5 or 6 percentage points—is sensitive to the definition of control total. The differences could potentially influence comparisons of top income shares within and across countries at a point in time as well over time.[43]

8.4.2.2.3 Definition of Top Income

As discussed above, the income definition follows the administrative requirements for tax statistics, which vary over time and across countries. In particular, the income definition used within the WTID attempts to be as close as possible to the definition of gross total market income (net of government transfers, taxes, and deductions). Changes in tax legislation may allow the inclusion or the exclusion of particular income sources within the reported income (e.g., capital gains, dividends, income deductions). In other words, these changes may bring about an expansion or a reduction of the tax base. We discuss below three specific types of changes and structural breaks in taxation regimes that can create severe problems for the consistency of top shares estimates over time. Nonetheless,

[42] Personal income minus transfers.

[43] There are, however, ways of avoiding the problem of control total altogether. The phenomenon of increasing inequality can be explained on one hand with the redistribution of resources toward the top 1% from the rest of the population (increase in top shares). On the other hand, this has proceeded together with a redistribution of income in favor of the very rich within the top percentile. The latter increase in inequality within the top is described by the Pareto coefficients or by the so-called shares within shares. These variables are, by construction, independent from the external control total.

we also point out that these changes do not always result in actual breaks (in levels or trends) for the top income shares series.

The first type of change in taxation discussed here deals with the treatment of deductions within the tax statistics. Starting from 1976, the income of the U.K. series, for instance, is grossed to include deductions that were previously subtracted from income[44]: "(i) allowable interest payments such as those for house purchase, (ii) alimony and maintenance payments, (iii) retirement annuity premiums, and (iv) other allowable annual payments" (Atkinson and Salverda, 2005). Such a change did not, however, cause a substantial change in the top income share: "the share of the top 1% was shown as rising from 5.6 to 5.7%, and that of the top 10% from 25.8% to 26.2%" (Atkinson and Salverda, 2005).

The second relevant type of taxation change concerns the treatment of capital income within the tax base. This problem is listed by Atkinson et al. (2011) as probably the "main shortcoming" of the WTID data, undermining the comparability of the top income series. Indeed, the estimation of top income shares is based on the observation of reported income for taxation purposes, and the restriction or the expansion of the tax base may be misleading representation of the real changes of total income held by top income groups.

On one hand, many sources of income from capital (interest income, returns on pension funds, imputed rents, etc.) have disappeared from the income tax base over time because they have been either fully exempted from taxation or are taxed separately. As reported by Iwamoto et al. (1995) and Moriguchi and Saez (2008), a substantial share of capital income, for example, was missing from the Japanese self-assessed income tax starting from 1947 "because almost all interest income has been either tax exempted or taxed separately and withheld at source . . . and so was a large part of dividends since 1965." However, as suggested by Moriguchi and Saez (2008), interests and dividends constitute only approximately 3% of total personal income in Japan, and even assuming that top groups absorb the whole income from these sources, the top 1% would still be far below the pre-1945 levels and below top 1% share in the United States. Similarly, the French tax base shrank with the exclusion of imputed rents of homeowners, as documented by Piketty (2001, 2003), who also provides some conservative estimates showing that the reduction of French top income shares was robust to the full imputation of tax-exempted capital income to reported income at the top.

On the other hand, the tax base may be expanding, generating a problem similar to, but the reverse of, the one discussed above. This is, for instance, described by Burkhauser et al. (2013) for the case of Australia, where the tax reform proposed in 1985 (and formally approved by 1987) aimed at broadening the tax base "in order to improve equity and efficiency."

[44] The income can be adjusted at the source for those countries where the information on deductions is available for income level. This is the case for the United States, as shown by Piketty and Saez (2006).

Most important, the tax reform included realized capital gains within the personal income tax base because "prior to 1985, Australia had no general tax on capital gains" and reduced substantially the marginal tax rates on dividends by introducing the so-called full-imputation system, which no longer allowed dividends to be subjected to both corporate and income taxation.[45] More specifically, both tax interventions (approved in 1987 and 1986, respectively), on the one hand, allowed the inclusion within the income tax base "most realized capital gains regardless of how long the asset was held. But to soften its effect, the reform applied only to assets purchased after September 19, 1985. Certain types of assets continued to be exempt, most importantly owner-occupied housing" (p. 8). On the other hand, the switch to a full-imputation system increases enormously, although artificially, the reported dividends income.[46] However, Burkhauser et al. (2013) further note that whereas the change in the tax law on dividends may have had an impact on the level of the share, the change in capital gains taxation had instead an impact that "grows over time with the stock of assets purchased after September 19, 1985 and the share of realized capital gains that enter the tax base" (p. 9). As reported by Burkhauser et al. (2013), these issues were not directly addressed by Atkinson and Leigh (2007) and led them to overstate the real increasing trend of Australian top shares.[47]

Finally, the third type of taxation change relates to the treatment of capital gains within the income definition. This can be more problematic because this source of income is particularly important for the very top income brackets—and is increasingly so because capital gains have been receiving advantageous tax treatment with respect to dividend-type income in most advanced countries. As distributed, corporate profits became less advantageous (dividends often are taxed at the income tax rate and subject to double taxation at the corporate level and individual level). Thus, including capital gains becomes fundamental "to assess the impact of retained profits of corporations on top individual incomes" (Atkinson et al., 2011). Moreover, because of favorable taxation, investors may be more willing to hold stocks with an underlying low payout ratio to cash in capital gains rather than dividends (e.g., the clientele effect). These considerations suggest that excluding capital gains can leave out a considerable (and increasing) amount of the income of richer tax units, making static and temporal comparison of effective top income shares across countries more problematic, assuming that the extent of the relevance of capital gains and their dynamics differ among countries.

[45] The Australian dividends tax reform in 1987 increased the corporate profits tax rate to the level of top income marginal tax rate (from 46% to 49%). However, as noted by Burkhauser et al. (2013), "under the new 100 per cent imputation tax system contained in the reform legislation, these company taxes effectively became withholding taxes. This was the case since their payment could be used to offset personal income tax on dividends as well as other taxes."

[46] To better understand the mechanical increase in reported income resulting from the switch to a full imputation system, we refer the interested reader to Burkhauser et al. (2013).

[47] This might be true, especially given the fact that Atkinson and Leigh's series include capital gains.

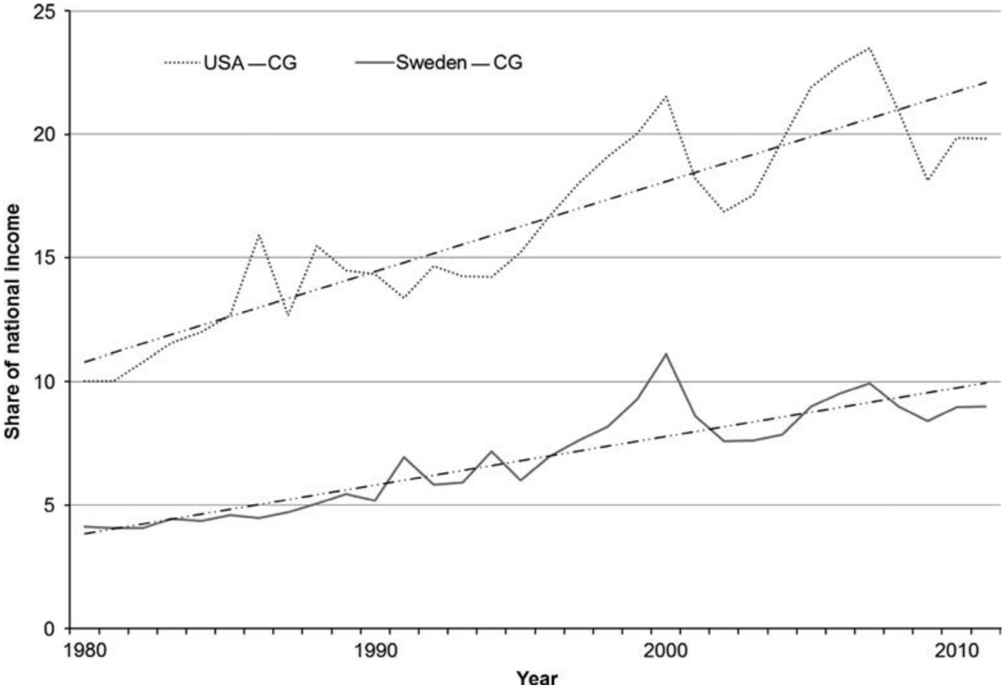


Figure 8.19 Top 1% share trends in the United States and Sweden, including capital gains (CGs). The graph shows that, despite the inclusion of CGs, the levels of the shares are substantially different across the two countries, whereas the trend over time becomes similar. *Source: World Top Income Database, accessed August 2013.*

To illustrate the validity of the argument above, we describe below how the top income share changes once capital gains are taken into account. However, this exercise can be done only for six countries, namely Canada, Japan, Germany, Spain, Sweden, and the United States.[48]

The role of capital gains for top income shares has been discussed by Roine and Waldenströmm (2012) for the case of Sweden, where, they argue, excluding capital gains "severely underestimates the actual increase in inequality and, in particular, top income shares during recent decades." Indeed, Figure 8.19 shows that after including capital gains, the top income share in Sweden has a similar trend to the top 1% in the United States. Yet the difference in level remains substantial. Figure 8.20 depicts the dynamics of top 1% income shares including and excluding capital gains for those countries for which data exist, suggesting that the importance of capital gains may also vary a great deal across

[48] Indeed, most of the top shares series within the WTID exclude capital gains altogether, whereas in a few countries capital gains are only included where taxable. The latter, however, cannot be untangled from the total income because no income source decomposition is provided. This is the case for the United Kingdom (before the introduction of a separate capital gains tax), Australia, New Zealand, and Norway.

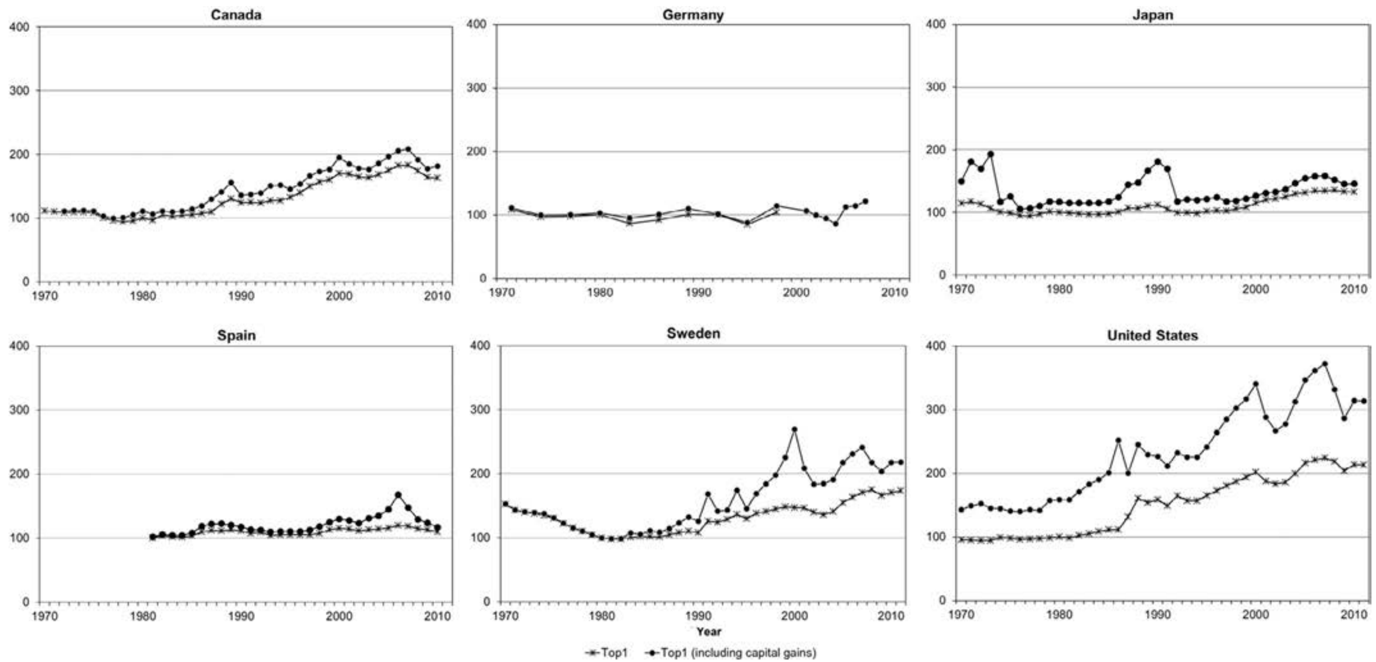


Figure 8.20 Cross-country variation in the impact of including capital gains income for top 1% share trends. The series of top 1% share excluding capital income are set equal to 100 in 1980. The series of top 1% shares including capital gains are calculated as follows: top1cg = 100 * top1cg/top1. *Source: World Top Income Database, accessed September 2013.*

countries. In the case of Germany, including capital gains income has essentially no impact on the top 1% share of income. Capital gains seem to affect the cyclicality of top shares in Japan, creating spikes but not persistent changes in level or trend. In Sweden, Canada, and Spain, the inclusion of capital gains income did not have a marked impact on the top 1% share before 1980 or 1990, but the influence of capital gains has become increasingly large since the 1980s, influencing the perceived trend of the increase in top shares. For the United States, the inclusion of capital gains income has resulted in systematically higher top shares over the entire period. It is crucial to note that the relevant concept of capital gains discussed here is that concerning its "realized" component for tax purposes. Indeed, realized capital gains refers to the wedge between the selling and purchasing prices of the asset. Furthermore, realized losses are subtracted from realized gains to obtain the measure of *net* realized capital gains valid for taxation purposes. This is a concept very different from accrued capital gains, which simply reflects the current differential between the "market price" and the purchase price. Indeed, individuals may realize capital losses for taxation purposes so that a series including capital gains is not necessarily more valid, informative, or complete than a series excluding capital gains.

Changes in tax systems can also affect the unit of reference (e.g., change in tax units); the extent of tax evasion and avoidance, including the phenomena of income shifting (e.g., substitute wages with tax-exempted noncash compensation); and anticipation or postponement of income returns. These issues are of crucial importance and are discussed later in separate sections.

8.4.2.2.4 Changes in Tax Avoidance and Tax Evasion

This section discusses the roles of (unlawful) tax evasion, (lawful) tax avoidance, and other behavioral responses to changes in taxation.

The use of tax data to estimate top income shares poses potentially serious problems resulting from underreporting, retiming of income reporting, and income shifting (depending on fiscal convenience). We discuss these important issues here to understand how they may affect the comparability of top shares series over time and across countries.[49]

Work by Piketty et al. (2012) showed that most of the countries under investigation experienced a reduction in the top marginal tax rate, which was highly correlated with the surge in top income shares we observed in the three decades following the end of the Bretton Woods system (see Figure 8.21). The reduction in the top marginal tax rate could indeed reduce the propensity to evade and avoid taxation, increasing tax collection and

[49] The specific composition of total reported income (e.g., capital vs. wage) may also be driven by tax convenience. In other words, capital and wage incomes are, to some extent, fungible and interchangeable. However, this issue is not discussed here because we are ready to assume that the "fungibility" of income sources does not affect the total reported income, only its composition. This issue becomes relevant in section 8.4.2.5 where we explicitly discuss the composition of income at the top.

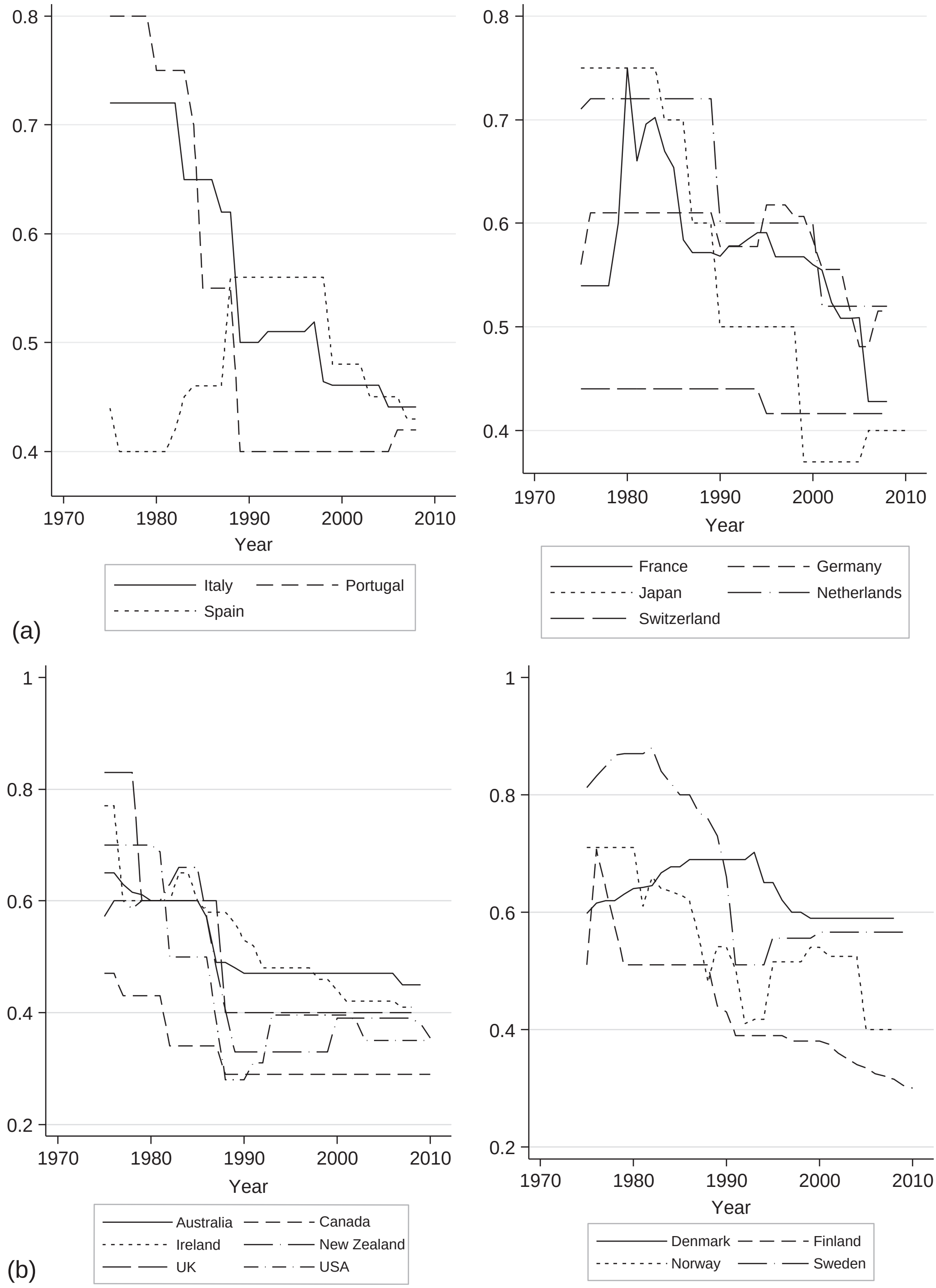


Figure 8.21 Top marginal tax rates across countries, 1970–2010: southern and continental European countries (and Japan) (a) and English-speaking and Nordic European countries (b). The figure depicts the top income tax rates (including both central and local government individual income taxes) over the period from 1970 to 2010. *Source:* Piketty et al. (2012).

therefore income reported at the top. Hence, the increase in inequality may be due to a reduction in tax avoidance because of lower tax rates for richer groups.[50] Indeed, Figure 8.21 shows how the top marginal tax rates changed over time for the sample of countries under investigation, highlighting a clear overall reduction in tax progressivity over time.

However, several researchers have devoted substantial attention to this aspect and pointed out that differences in levels across countries and the upward trend in income inequality observed in many countries are substantially robust, real phenomena rather than spurious results merely driven by tax avoidance and tax evasion (Alvaredo, 2010; Alvaredo and Saez, 2009; Banerjee and Piketty, 2005; Leigh, 2009; Leigh and van der Eng, 2009; Moriguchi and Saez, 2008; Roine and Waldenström, 2008).

First, as reported by Leigh (2009), the evidence suggests that income underreporting (the size of the tax gap) does not substantially vary across countries, whereas tax regimes vary dramatically. In addition, the extent of tax avoidance (and the scope for evasion) at the very top of the income distribution may not necessarily be higher than that for the rest of the distribution given the public visibility of their sources of income and the efficient enforcement efforts of tax authorities (Alvaredo, 2010; Alvaredo and Saez, 2009; Leigh and van der Eng, 2009).

Second, despite the reduction in tax progressivity, one can argue that there is little evidence to suggest that the extent of tax avoidance among the richest households has changed substantially over time, at least for countries with relatively high tax compliance[51] (see Internal Revenue Service, 1996, 2006 for evidence on the United States; Roine and Waldenström, 2008[52] for Sweden). In addition, even if this was not the case, the extent of avoidance (and evasion) at the top should have decreased much more than that of the rest of the population to exert an overall positive influence on the top shares.[53]

Third, the country-specific top wage shares often closely follow the evolution of the overall top income shares. These results are inconsistent with the assumption that the evolution of inequality is entirely captured by the time-varying tax avoidance and evasion. Indeed, the tax on wages and salaries—compared with farm income or business income—is usually withheld at the source so that it is almost impossible to escape the tax authorities' purview (Alvaredo, 2010; Banerjee and Piketty, 2005; Moriguchi and Saez, 2008).

[50] This case is made by Reynolds (2007) for the United States.

[51] This might be because government and tax authorities already had strong incentive and capability to enforce tax regulations in place when overall marginal rates at the top began to be reduced. Nonetheless, no evidence is available for countries outside the United States and Sweden, and any further generalization is not prudent.

[52] The authors suggest that it is possible that both the incentive of top individuals to underreport and the benefit and ability to monitor tax compliance by tax administration went up contemporaneously.

[53] A similar argument is used by Williamson and Lindert (1980).

Finally, Piketty et al. (2012) developed and tested a model linking the increase in top income shares to reduced tax progressivity, mainly through three motivations: increase in hours of work supplied (e.g., supply-side theory), increased rent-seeking activity (e.g., a top executive has control of salaries has more incentive to influence remuneration and seize a greater share of the firm profits), and decreased tax avoidance. However, the elasticity of reported income to change in tax rates due to tax avoidance is considered the least important factor. Indeed, Piketty et al. also point out that there are countries (such as Italy, Japan, Sweden, and the Netherlands) that experienced only modest increases in top shares despite significant top rate tax cuts of similar magnitude to those implemented in Norway, Finland, and all advanced English-speaking countries.

In sum, there is little evidence to suggest that tax avoidance in relation to decreases in top marginal tax rates is a particularly relevant explanatory factor of the long-run surge in top income shares. Nonetheless, other types of changes in taxation regulations may well have a substantial impact on top shares. Below, we differentiate between those changes bringing about permanent shifts in income or temporary behavioral responses.

Both theory and empirical evidence highlight how tax avoidance (especially income shifting over time or across the tax base) can be of great relevance for short-term changes in reported income to exploit tax opportunities (Saez et al., 2012 provide a comprehensive survey of this literature as well as interesting empirical findings). The classic example discussed within the top incomes literature is the Tax Reform Act (TRA) of 1986, which, among other things, dramatically decreased the top marginal tax rate on personal income, increased the capital gains tax rate, and, while lowering it, set the corporate income tax rate at a level higher than the top personal income tax rate. These policy changes provided strong incentive for agents to realize capital gains in the short term and to shift income from corporate to personal income (Gordon and Slemrod, 2000; Slemrod, 1996). It is worth noting that the shifting of business income from the corporate tax base to the individual tax base has brought about a permanent level shift for the top income shares in the United States (if excluding capital gains). Following these legal changes, the highest fractile share in total U.S. income (top 0.01%) increased by 30% from 1986 to 1987 (a year of a stock market crash) and by 53% from 1987 to 1988 (when a systemic banking crisis hit). However, as argued by Atkinson et al. (2011), the taxation policy change did not affect the series including capital gains[54] because "before TRA 1986, small corporations retained earnings and profits accrued to shareholders as capital gains eventually realized and reported on individual tax returns. Therefore, income including capital gains does not display a discontinuity around TRA 1986" (note to Figure 8.5). Similarly, the changes in taxation do not seem to have influenced the long-run trend of the shares.

[54] The only documented effect was an artificial spike in capital gains realizations in 1986 in anticipation of the announced increase in tax rate that took place in 1987.

There are also other less-discussed examples that may illustrate the full range of changes (and their complexity) in tax statistics and that we ought to take into account. For example, a substantial reduction of marginal tax rates on dividends was obtained in Australia in 1987 through the introduction of the full-imputation system (before the income from dividends was subjected to both corporate and income taxation).[55] Atkinson and Leigh (2007) note that "the effect of the introduction of imputation in Australia in 1987 is evident in the statistics." Indeed, the taxation regime change was announced in 1985 and had a short-term impact on dividend distribution once the law was passed in 1987, consistent with optimal retiming of income reporting. Moreover, it is also possible that the substantial reduction of the marginal tax rate on dividends may have induced more firms to distribute a greater share of their profits (changing the level of the shares).[56]

A switch to a full-imputation system in 1993 had a long-lasting impact on the composition of top income shares in Finland (as documented by Jäntti et al., 2010).[57] A similar change was documented for New Zealand in 1989. However, in the same year, a tax cut to take place in 1990 was announced (the top individual rate would have been reduced to the company tax rate), causing companies to postpone their payment to top executives. Moreover, "similar anticipation of tax changes is likely to have caused the sharp spike in top income shares is observed in 1998–99, and may have caused the 2000 figure to be depressed," as discussed by Atkinson and Leigh (2005).[58] Similarly, in 2005 Norway announced a permanent increase in dividend tax (to be increased in 2006); this marked

[55] The Australian dividends tax reform in 1987 increased the corporate profits tax rate from 46% to 49%, equalizing it to the top income marginal tax rate. However, as noted by Burkhauser et al. (2013), "under the new 100 per cent imputation tax system contained in the reform legislation, these company taxes effectively became withholding taxes. This was the case since their payment could be used to offset personal income tax on dividends as well as other taxes."

[56] It is nonetheless important to note that, as argued in recent work by Burkhauser et al. (2013), the changes in taxation regime also have brought about substantial permanent effects on reported income within top income brackets, although not through taxation avoidance. The previous section discussed how the taxation reform expanded the personal income tax base through the inclusion of dividends and capital gains (possibly more gradually), which were previously unreported. Once these issues are appropriately taken into account, the real increasing trend of Australian top shares is slightly downsized, especially using the series including capital gains. Indeed, as discussed earlier, it may be possible that the inclusion of capital gains within the personal tax base was only gradually increasing over time given the specific prescriptions of the taxation reform.

[57] The introduction of a dual taxation system that favored capital income promoted the surge of dividends income at the expense of entrepreneurial income.

[58] The authors go on to state that when the government announced in 1998 that "the marginal tax rate on earnings over $60,000 would be raised from 33 percent to 39 percent in the 2000 tax year, many taxpayers took the opportunity to realise business earnings in the 1999 tax year, significantly boosting top income shares in that year, and perhaps to a lesser extent also in the 1998 tax year."

a notable peak in top income shares as individuals and corporations shifted income over time to avoid the impending rate increase (Aaberge and Atkinson, 2008).[59]

8.4.2.2.5 Definition of Tax Units

Different country-specific series are based on different definitions of tax units. In some countries the unit of reference is the family (e.g., typically spouses with dependents or singles with no dependents). This is the case for the United States and most continental European countries. Other countries define their tax units based on individuals. This is the case, for example, for Australia, Canada, New Zealand, Japan, India, Italy, and Spain.[60]

Most important, some countries have experienced a change in the tax base as the taxation system moved from a family to an individual base. Fortunately, however, only the United Kingdom experienced such a shift within the period under analysis (the shift occurred in 1990).[61] Such a change in tax units can create comparison problems for at least two reasons.[62]

First, the level of top shares is affected and the direction and magnitude of such a change depends respectively on the joint distribution of income within families populating the top income brackets and on the actual proportional difference between the number of individuals and the number of tax units as well as on the specific assumption about the Pareto coefficient.[63] As discussed by Atkinson et al. (2011), if the income for the richest families is unequally distributed (e.g., the head of the family concentrates most of the family income), we expect, under specific assumptions, the shift from family to individual unit to have a positive impact on the measured top income share series. The impact on the shares becomes negative if income is equally distributed within the top tax units. In the United Kingdom, for instance, the tax units increased from 33,000 to 46,000 in 1990 because of the change from family units to individual units, and the top 1% series experienced an approximate positive jump of 1 percentage point as a result.

[59] Aaberge and Atkinson (2008) also proposed a new set of estimates of Norwegian top shares by estimating the capital income from stock holding using a "Hicksian" approach. In other words, they impute the returns from stocks by multiplying the "estimated market value of the households' stocks and the long-run average rate of return (8.9 per cent) on the Oslo Stock Exchange (OSE)" (p. 13). Moreover, they argue that "The 'Hicksian' measurement of the stock returns is less sensitive to changes in income reporting behavior than the conventional income definition and may thus provide a better basis for analyzing the trend in top incomes during the pre- and post-reform period."

[60] The unit of analysis in New Zealand was based on family before 1953.

[61] Also, Spain changed from family to individual taxation in 1988, but this was corrected within the original calculations of the series. See Alvaredo and Saez (2009). Other cases of changes in tax units occurred in the pre-1970 period are discussed within Chapter 7. In particular, the authors discuss the "borderline" case of Sweden, in which, "the family was the tax unit before 1967 when a choice of filing individually was introduced. This was then the rule until individual taxation finally became compulsory in 1971" (p. 17).

[62] We do not discuss explicitly the change in the control total for the population that has to change accordingly to the definition of the tax units.

[63] Atkinson and Harrison (1978, in particular, Chapter 9).

Second, and most important, a change in the composition of tax units may also affect the trend of the series, not just its level. Indeed, this happens if the factors influencing the level of the shares discussed above also vary over time. This is not implausible; for example, income may have become more evenly distributed within the richest families and the growth rate of tax units could have well exceeded population growth over the past decades. Also, the change in the distribution of income at the top over time (formalized as the change of the Pareto coefficient) has been thoroughly documented and discussed by Atkinson et al. (2011). Nonetheless, the available country-specific evidence (for Canada) shows that the use of different unit bases may affect only the level of shares (see Saez and Veall, 2005).

8.4.2.2.6 Gross and Disposable Top Income Shares

Gross income data can be complemented by information on government transfers and taxation to obtain measures of disparity in disposable incomes or spendable income, ultimately a preferable income definition for individuals. Indeed, the pretax top share can show a different picture than the post-tax share, depending on the degree of progressivity of the tax system and the extent of redistribution. As discussed earlier, tax systems in most of the countries discussed here have changed a great deal over time and have reduced their progressivity. These tax policy changes can influence both the perception of economic inequality and comparisons of inequality over time and across countries. Indeed, the incidence of taxation on net top income shares may vary across countries, affecting the extent of comparability of top shares trends across countries.

Data on disposable top income share are available only for a handful of countries. In this section we describe the evidence for the Netherlands and the United Kingdom (Atkinson and Salverda, 2005), Canada (Veall, 2012), the United States, and France (Piketty and Saez, 2006).

Although one should bear in mind that methodologies adopted by these authors are not homogenous and income definitions are not directly comparable, it is interesting to obtain a measure of the direct impact of taxation on top shares.

Following the work by Atkinson and Salverda (2005), we divide the pretax income shares by the after-tax shares to measure the so-called relative implicit tax rate. We define the latter as the "arithmetic impact of taxation" on top shares, calculated as 1 − (pretax share)/(post-tax share), which in turn is equal to 1 − (1 − average tax rate at the top)/(1 − average tax rate for the overall population). Figure 8.22 depicts the "implicit tax rate" for Canada, the Netherlands, the United States, and the United Kingdom,[64]

[64] France has only two observations, one in 1970 and one in 2005, which are not shown in the graph. The implicit tax rate calculated for the French top 1% share is relatively mild and it marginally increased over time—from 0.08 in 1970 to 0.1 in 2005 (results are based on calculations of data taken from Piketty and Saez, 2006). In other words, the net of taxes top 1% share is approximately 8% and 10% lower than the gross share in 1970 and 2005, respectively). We remind the reader here that the calculated implicit tax rates are not directly comparable across countries.

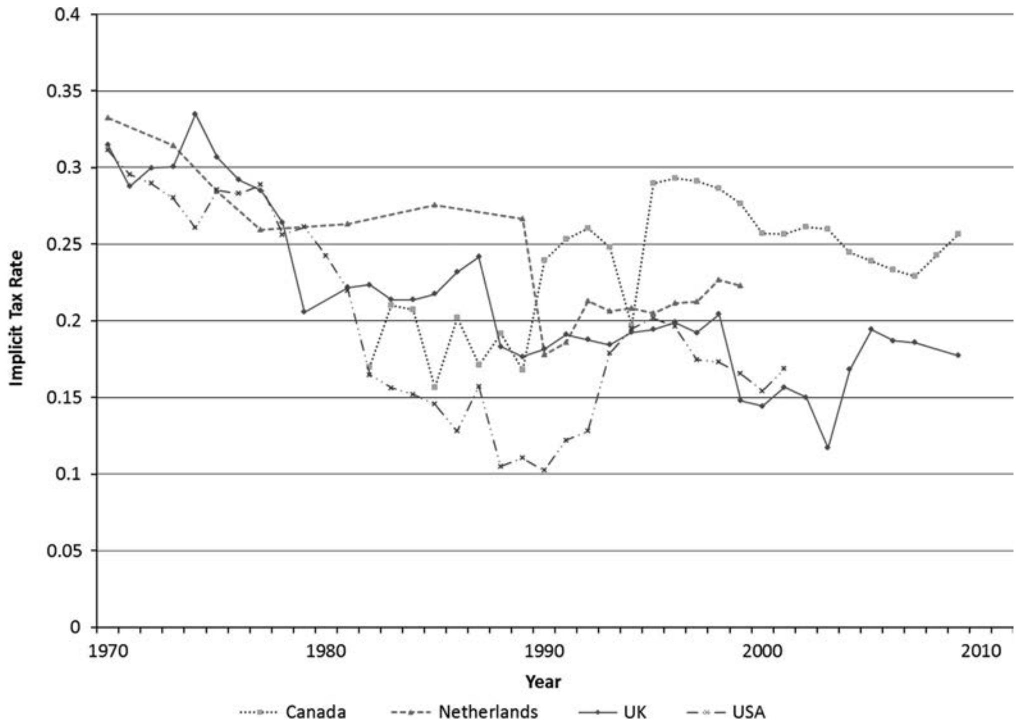


Figure 8.22 Implicit tax rates for a selected group of countries. The graph shows the dynamics of the "relative implicit tax rate" from 1970 for the United States, United Kingdom, Canada, and New Zealand. The implicit tax rate represents the "arithmetic impact of taxation" on top shares, and it is calculated as [1 − (pretax share)/(post-tax share)], which in turn is equal to [1 − (1 − average tax rate at the top)/(1 − average tax rate for the overall population)]. *Sources: Calculation of the authors based on data from country-specific literature for the United Kingdom and New Zealand (Atkinson and Salverda, 2005), Canada (Veall, 2012), and the United States and France (Piketty and Saez, 2006).*

showing that, with the exception of Canada, the tax system reduced its progressivity in these countries. For the United Kingdom, United States, and the Netherlands, the inversed implicit tax rate for the top 1% went from around 35% in 1970 to around 20% in 2000. However, the implicit tax rate in United States decreased more during the 1980s, reaching the value of 10% in 1990 before rebounding to around 20%. In Canada the pattern was nearly reversed. The Canadian implicit tax rate was, on average, lower than 20% during the 1980s; it then increased by 10 percentage points during the 1990s and finally declined gradually to a value of around 25%. This is why net top income share would show an attenuated increase in inequality in Canada, as shown in Figure 8.23. In the United States and the United Kingdom, however, pre- and post-tax trends in the top 1% share are essentially indistinguishable.

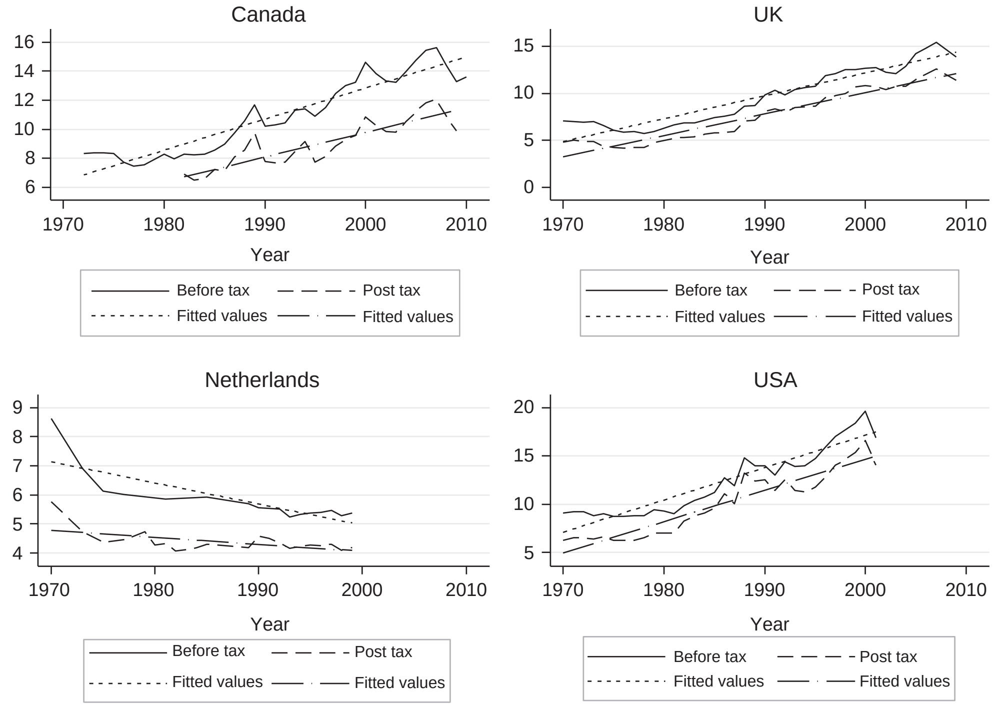


Figure 8.23 Pre- and post-tax top 1% shares for selected countries. The graphs show both the top 1% based on gross income (net of taxes and of transfers) as well as on net income. *Sources: Elaboration of the authors based on data from country-specific literature for the United Kingdom and New Zealand (Atkinson and Salverda, 2005), Canada (Veall, 2012), and the United States and France (Piketty and Saez, 2006).*

8.4.2.3 Top Shares in the Late 2000s

The WTID contains information for top income shares in 25 countries. In addition to all of the caveats described above, making comparisons across countries is limited to the years of data that are available. Nineteen of those countries, though, do have data during the "late 2000s" (2009, 2010, or 2011), and several more have at least some data for the period of the "mid-2000s" (2003–2008). Such data for the top 1% shares in the mid-2000s and late 2000s are represented in Figure 8.24. Comparisons of the level of inequality using these top share figures across countries may be problematic for all of the reasons discussed above. Differences in definitions of income and income reporting units, as well as tax treatment of different types of income and potential differences in tax reporting, avoidance, and evasion, can all influence differences in the levels of top income shares over time.

These caveats notwithstanding, it is interesting to notice that the ranking of countries based on top income shares remains similar to what was observed using the inequality

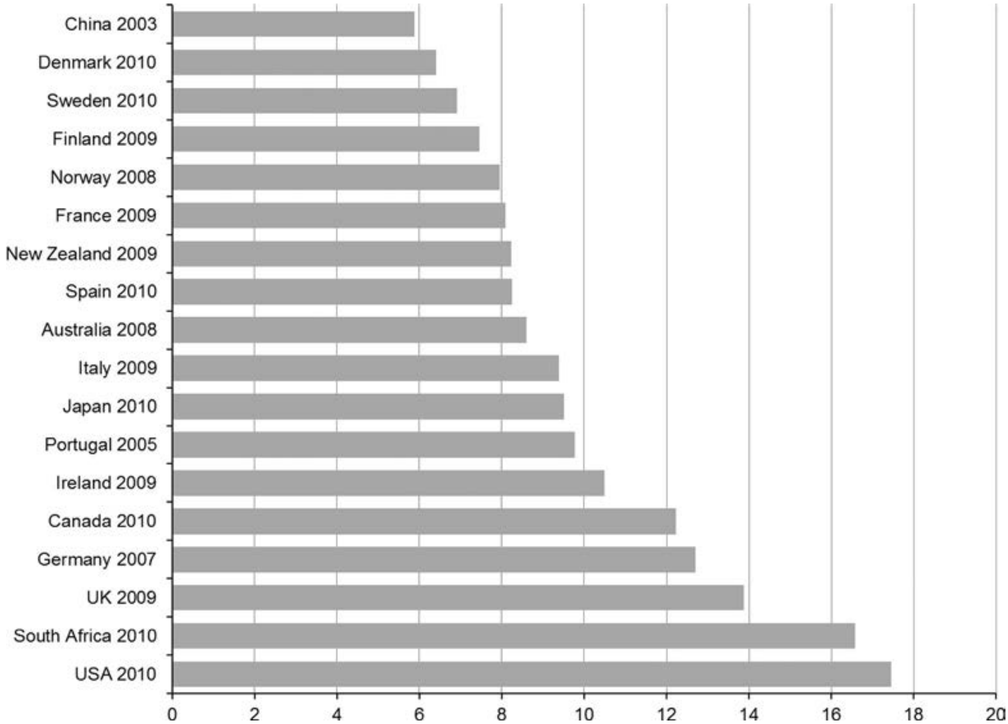


Figure 8.24 Top 1% shares in the late 2000s. *Source: Data from the World Top Income Database (accessed September 2013).*

measures across in the entire distribution (based on data comparable across countries) in Section 8.4.1. Among rich countries, the English-speaking countries have higher measured inequality than the Nordic counties. In 2010 the top 1% share was nearly 18% in the United States and less than 7% in Sweden and Denmark. Among the MICs and developing countries under investigation within the chapter and for which we have data, South Africa has the highest levels of inequality.

8.4.2.4 Comparison of Trends Across Country Groups

The primary goal of this section is to explore commonalities and differences in data trends across countries from 1970 to 2010. Previous sections emphasized how different sources, methods, and definitions of income tax may affect the estimated top shares across countries. However, we have showed that not all changes in methodology or breaks in data series create comparability problems.

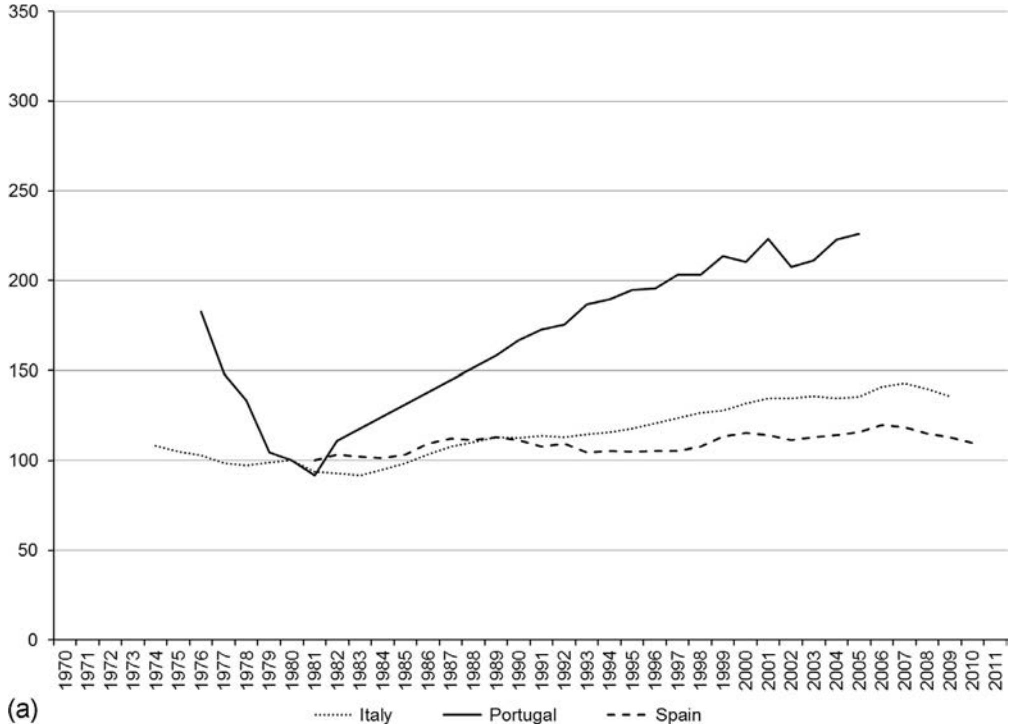
As noted by Gottschalk and Smeeding (2000), the time-varying and time-invariant factors specific to the shares need to be the same across countries to have meaningful cross-country comparisons. To ease the exploration of differences in trends, we ignore factors that are country-specific and time invariant, namely the differences in levels. More precisely, we standardize the values of the shares to be equal to 100 in 1980. Because of data availability, the standardization to 100 is done for the year 1990 for emerging countries. This takes care of measurement errors and heterogeneity of methodology of calculation of top shares across countries that are constant over time.

We group the countries in our data set into the following clusters: Nordic European (Denmark, Finland, Norway, and Sweden); southern European (Italy, Portugal, and Spain); western English-speaking (Australia, Canada, Ireland, New Zealand, the United Kingdom, and the United States); and continental European countries (France, Germany, the Netherlands, and Switzerland) together with Japan. The remaining countries—China, India, and South Africa—are labeled as emerging or MICs. These clusters differ somewhat from the previous section discussing inequality across the entire distribution but are consistent with the groupings used by Atkinson et al. (2011). According to Atkinson et al. (2011), these groupings are "made not only on cultural or geographical proximity but also on proximity of the historical evolution of top income shares" (p. 40).

The various panels in Figure 8.25 show that top income shares are growing in many countries; increasing inequality is not limited to a small number of countries or any obvious subset of countries. Indeed, a common pattern observed across most of the countries in the WTID shows declining top shares for one or two decades since 1970, followed by steadily rising top shares through 2010. The precise timing and extent of the "U turn" in top shares varies across countries, and we provide below a description of the main features of the dynamics of top shares over time across different country groups.

All of the southern European countries have seen an increasing top 1% share since 1980, but the increase has been much sharper in Portugal, where the top share more than doubled between 1980 and 2010 compared with an increase of "only" 40% in Italy and approximately 15% in Spain (Figure 8.25a). Trends in the top shares of Continental European countries (Figure 8.25b) fluctuate more with business cycle patterns than the southern European countries. Nonetheless, the pattern of top shares series remains broadly consistent with a mild U shape; the top 1% share fell between 1970 and the early 1980s and remained more or less stable until the mid-1990s, when it mildly rebounded until the onset of the 2007–2008 financial crisis. In fact, most of the Continental European countries lack top share data over the past decade, making a complete analysis of this period impossible at present. France and Japan are the two countries in this group with data over the full period, and they both follow this pattern closely. Between the mid-1990s and mid-2000s, the top share increased approximately 30% in Japan and 15% in France. Between 2007 and 2010 top shares held steady in Japan but declined sharply in France.

Top shares in the English-speaking countries also fluctuated with the business cycles but exhibit a clearer upward trend since the early or mid-1980s than the Continental European countries (Figure 8.25c). Moreover, with the exception of New Zealand,



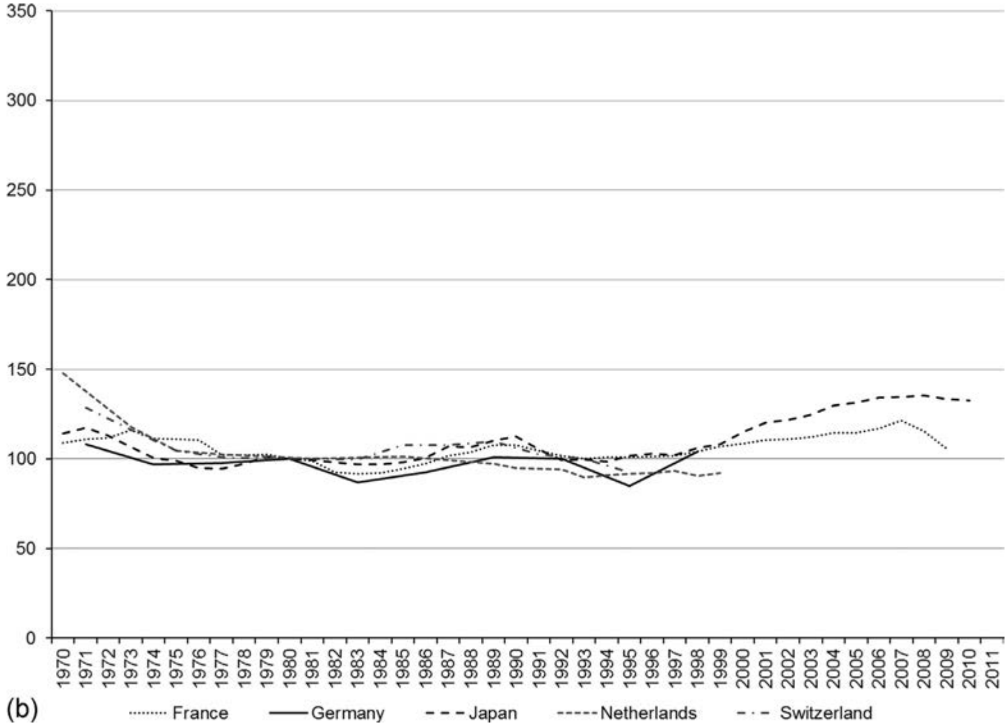
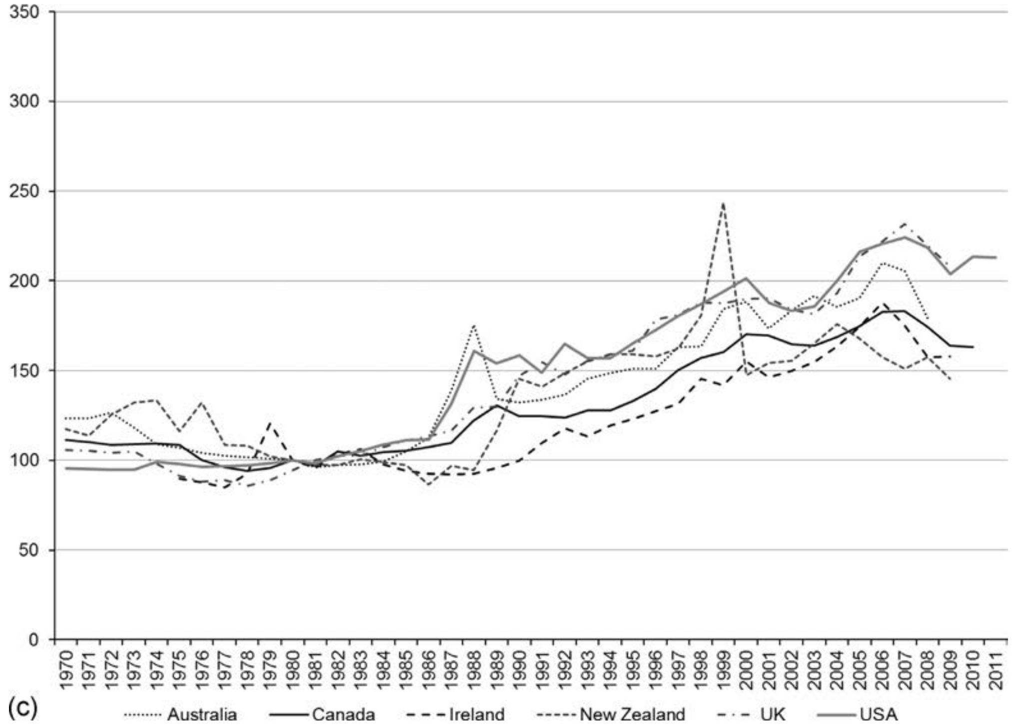


Figure 8.25 Top 1% share trends between 1970 and 2011 by country group: Southern Europe (1980 = 100) (a). Continental Europe and Japan (1980 = 100) (b), English-speaking countries (1980 = 100)

Continued



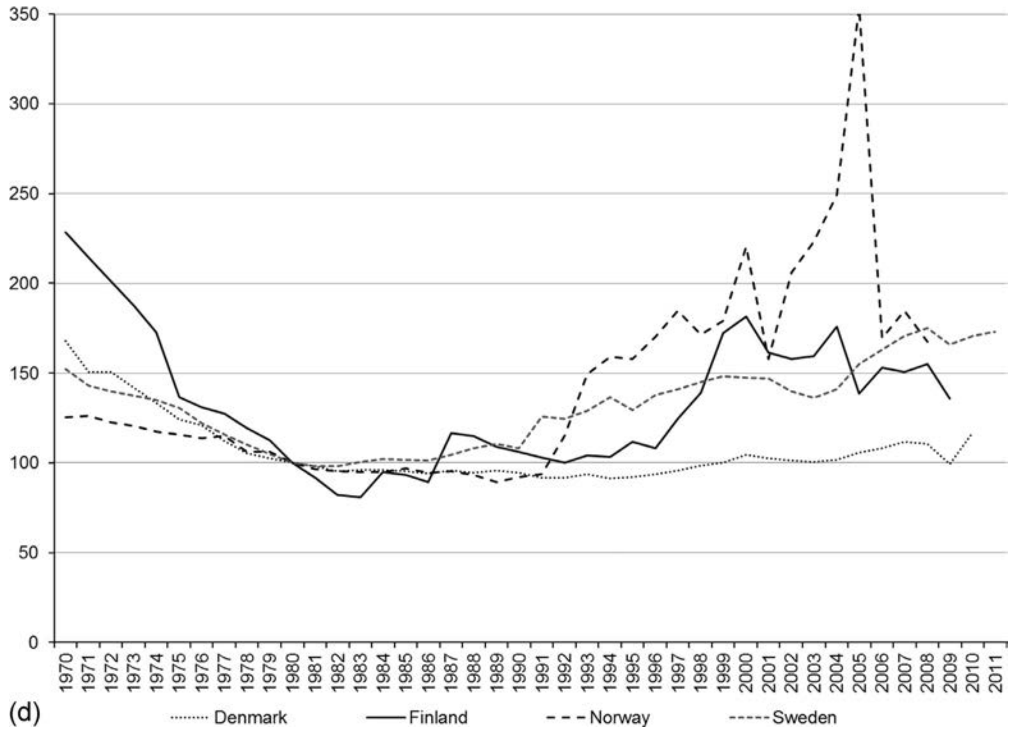


Figure 8.25, cont'd (c), Nordic countries (1980 = 100) (d), and developing countries (1990 = 100)

Continued

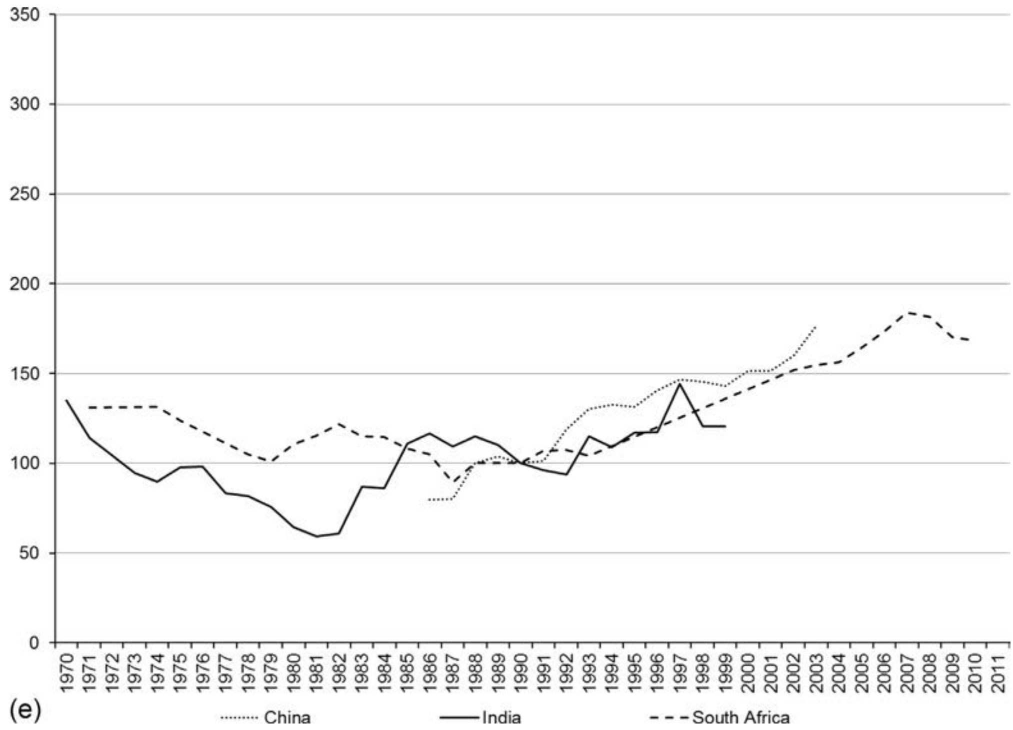


Figure 8.25, cont'd (e). In (a)–(d), 1980 was calculated as 100. In (e), 1990 was calculated as 100. *Source: World Top Income Database, accessed September 2013. Elaboration by the authors.*

all English-speaking countries[65] experienced a similar trend since the end of the 1980s.[66] Between 1990 and the onset of the 2007 financial crisis, the top 1% share increased between 60% and 70% (in Australia, Canada, the United Kingdom, and the United States) and around 90% (in Ireland). Trends in the English-speaking countries show some evidence of the impact of the economic crisis, with top shares decreasing between 2007 and 2010–2011.

[65] Evidence for New Zealand suggests that the top 1% share increased by 50% from 1980 to the onset of the recent crisis, accounting for only the change in reported income driven by the switch to the full imputation system in 1989, as detailed earlier. Once we ignore this permanent jump in reported income (and the subsequent temporary spike in 1998–1999 due to income shifting over time), the trend for New Zealand's top share is almost flat. This was already noted by Atkinson and Leigh (2005).

[66] This allows netting out some of the structural effect induced by changes in the taxation system, as discussed earlier.

In the Nordic countries, top shares were mostly flat during the 1980s and did not start increasing until 1990 or later (Figure 8.25d). This is particularly clear in the case of Norway, where the top 1% share was unchanged between 1980 and 1990 but doubled between 1990 and 2000. Increases after 1990 were smaller in the other Nordic countries, especially Denmark, where the top income share only rose 15% between the late 1980s and the late 1990s before sliding back down in the late 1990s and early 2000s. The post-1990 trend in rising top shares appears to have been halted or reversed by 2000 in Finland and by the mid-2000s in Norway, although in Sweden the increase in top shares continued. As discussed earlier, the unusually large spike Norway's top 1% share in 2005 is attributable to dividends paid out in anticipation of tax policy changes in 2006.

In developing countries in the WTID, the trends in top shares seem to resemble most closely those in the English-speaking countries. Top shares started increasing in the early 1980s in India and in the late 1980s in China and South Africa (Figure 8.25e). After the 1990s, these three developing countries appear to experience a long-term increasing trend in the top 1% share of income.

Between 1980 and 2009–2011, top shares more than doubled in the United States and the United Kingdom and were on track to double in Australia and Ireland before falling sharply in the global economic downturn that hit most developed countries beginning in 2007/2008. The English-speaking countries stand out as experiencing the largest increases in top shares over the entire post-1980 period, and they account for three of the top five countries with the largest cumulative changes in top 1% share between 1980 and the average after 2000 (Figure 8.26a). However, much of the run-up in top shares in English-speaking countries occurred in the 1980s. Focusing on changes in top shares since 1990, a different set of countries stands out as having large increases. The four countries with the largest cumulative changes in the top 1% share since 1990 include two Nordic countries (Finland and Norway) and two developing countries (China and South Africa) (Figure 8.26b). After 1990, cumulative increases in top shares were roughly equal between English-speaking and developing countries. It is important to note that the results are invariant to the exclusion of the abnormal spike in the top income share that occurred in Norway largely as a result of the anticipated change in taxes on dividends.[67]

The various graphs in Figure 8.25 track the decrease in top income shares in the decade or two after 1970, depending on the country, followed by increasing top shares starting around 1980 or 1990, again depending on the country. Whether these long-term

[67] As mentioned before, in 2005 Norway announced a permanent increase in dividend tax (to be increased in 2006); this marked a notable peak in top income shares as individuals and corporations shifted income over time to avoid the impending rate increase (Aaberge and Atkinson, 2008).

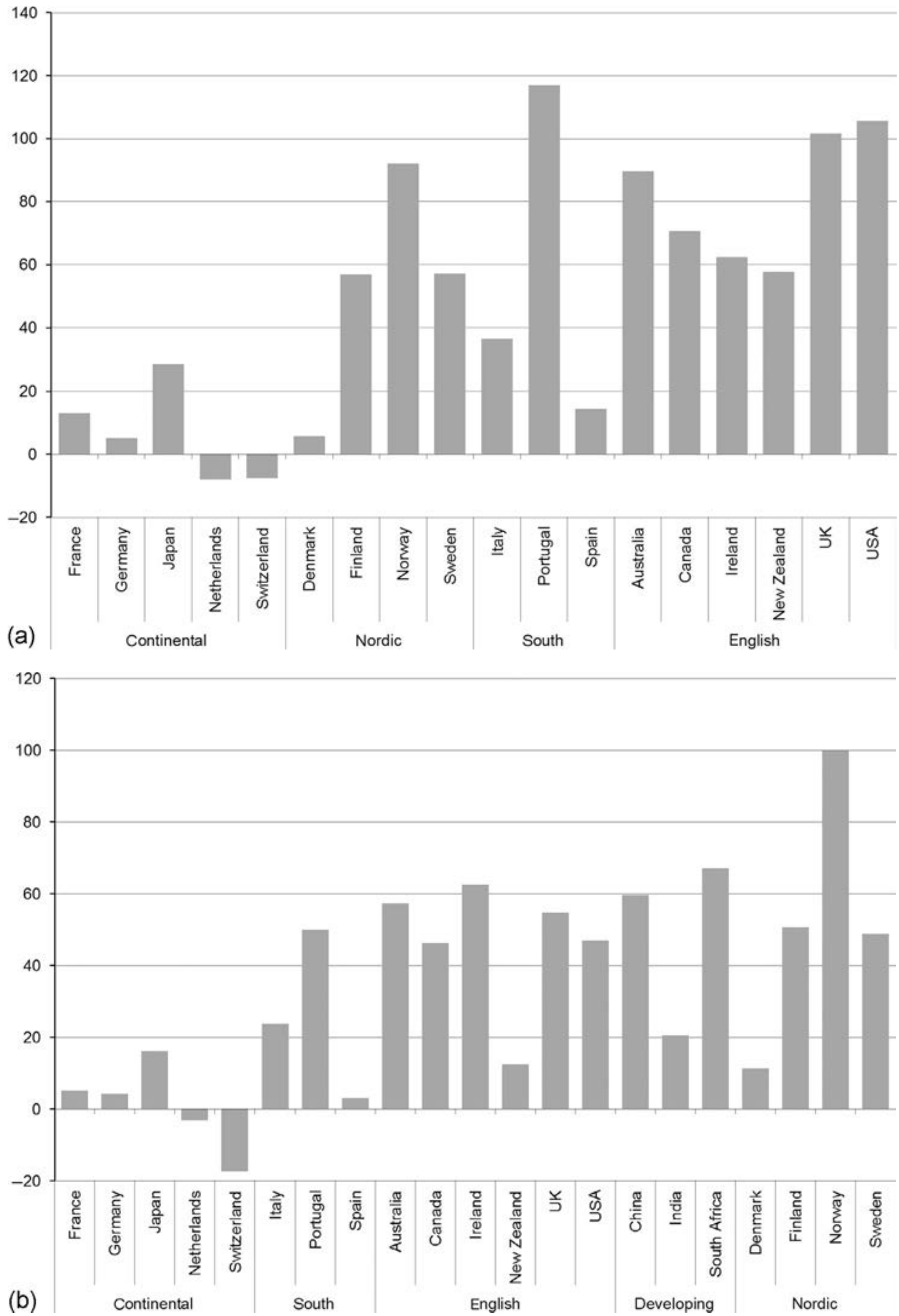


Figure 8.26 Cumulative changes in top income shares by country and country group. (a) Average cumulated change from 1980 to after 2000. (b) Average cumulated change from 1990 to after 2000. Data are sorted by the average cumulative change in top 1% by country groups. In the case of Germany, we draw information from the top share including capital gains. The cumulated change is computed since 1990 for the developing countries. Because of a lack of information the period after 2000 is equivalent to 1995 for Switzerland and to 1999 for India and the Netherlands. Results obtained exclude Norway's top 1% peak in 2004 and 2005. *Source: World Top Income Database, accessed September 2013. Elaboration by the authors.*

trends will persist into the future is an open question. Most of the countries in the WTID did witness decreasing top shares between 2007 and 2010, but many of the countries lack data during these years. Have long-term trends toward rising inequality been reversed by this period of financial turmoil and recession? Or, can they be expected to revert to their pre-2007 trends? Morelli (2014) uses the US top income shares data to answer this question by estimating impulse response functions of top shares to the occurrence of banking shocks. The main findings of this paper suggest that the short-term impact of systemic banking crises on the upper-income brackets of the income distribution is negative at the very top of the income distribution (e.g., above the 99th percentile) and positive at the bottom of the top decile (e.g., between the 90th and 99th percentiles). In other words, the relative response to systemic banking shocks differs across top income groups given their heterogeneous nature.

However, and most importantly, systemic banking crises do not seem to substantially affect top income shares; their estimated dynamic responses are found to be relatively small in magnitude.[68] Moreover, the findings of the paper are also suggesting that the impact of crises may also be temporary in nature since top income shares may quickly return to their predicted path in the absence of a crisis.

Consistent with what was informally documented and suggested by Atkinson et al. (2011), Saez (2013), and Piketty and Saez (2012), these results suggest that even major disruptive crises such as the financial turmoil in 2007/2008 do not represent a structural break for top shares series, and we should not expect a reversal of the increasing trend in income concentration unless some strong change in the political and institutional framework is expected (e.g., a change in taxation regime, remuneration practices, regulation policies).

8.4.2.5 Income Decomposition

In the sections above we described the increasing trends of top income shares for most of the countries. What is driving these trends? To better understand the mechanisms that led to an increase in inequality in most of the countries under investigation, we can use composition data from income tax statistics. However, income decomposition by sources is available for only a few countries. Ideally, we would like to understand how both marginal distributions of each source of income as well as their joint distribution affect the dynamics of the right tail of the income distribution. This is discussed by Atkinson et al. (2011) and subsequently by Alvaredo et al. (2013) in the case of the United States for two sources of income, namely wage and capital. Their results suggest an increasing

[68] Indeed, the estimated impacts of crisis are generally found to be substantially below one standard deviation, irrespective of the specific top share or country group under investigation.

association between the two sources of income for individuals within top brackets. However, understanding this important issue (which requires the availability of microdata for every country and year) goes well beyond the scope of this chapter. A less rigorous approach is to simply decompose the top income into, say, two main sources (wage and capital income) to understand their incidence in total income accruing to the top.

Below we depict the share of capital income (including rental income from buildings, interest income, and dividends but excluding realized capital gains where possible) and employment income (wages, salaries, bonuses, allowances, and pensions) for those eight countries where these calculations are possible (Australia, Canada, France, Japan, Italy, the Netherlands, Spain, and the United States). On balance, wage income weighs substantially more within the total top income above the 99th percentile (top 1%). This holds true with the exception of Italy and Australia, where wage income has a relatively lower incidence on the total (Figure 8.26).[69] The picture is reversed if we look at richer top income brackets above the 99.99th percentile (top 0.01%). Here, the incidence of capital income is generally higher than earned income (the only exceptions are Canada and the United States).[70] Results are shown in Figure 8.27.

The relative shares of different sources of income accruing to the top also have changed over time, and the experience has been heterogeneous across countries. In the case of top 1% shares, on balance, there is evidence of a slight increase over time in the labor-type income share for the countries for which income composition data are available (Figure 8.27a). The main exceptions are Spain and Australia, which exhibited decreasing wage shares from the late 1990s until the mid-2000s.[71] In Japan and the United States, wage shares increased slightly before 1990, but they have remained roughly constant since. In other countries, including France and Italy, there is little evidence of any trend

[69] Australian data can only be decomposed as wage and nonwage income. The latter represents our definition of "capital" income and includes realized capital gains as well as business income (self-employed income, profits from unincorporated businesses, and farm income).

[70] Recall that realized capital gains are not included in the standard top income shares under analysis.

[71] It is important to note that Spanish data include capital gains within the definition of capital income. This explains the more acyclical nature of wage shares, which recovered substantially in Spain in the late 2000s; a similar but milder trend seems to have occurred in Australia. It is also important to notice that Australia remains the only country where the incidence of capital income within the top 1% group is higher than the wage share. This is due to two main reasons. First, business income is entirely included within the "nonwage" income. Second, Australian data incorporate capital gains to the extent that they are taxable (at varying degrees over time). Indeed, as recalled by Burkhauser et al. (2013), "Prior to 1985, Australia had no general tax on capital gains. Hence, almost no capital gains were captured in tax record data, since most capital gains were excluded from the personal income tax base. It was not until 1972 that realized capital gains on most assets were included in the tax base under Section 26AAA of the tax law, but only for short-term capital gains (those on assets held less than one year), and excluding owner-occupied housing. While a tax on realized capital gains on assets based on speculation existed as far back as the 1920s, it was not systematically enforced and it generated little revenue" (p. 8).

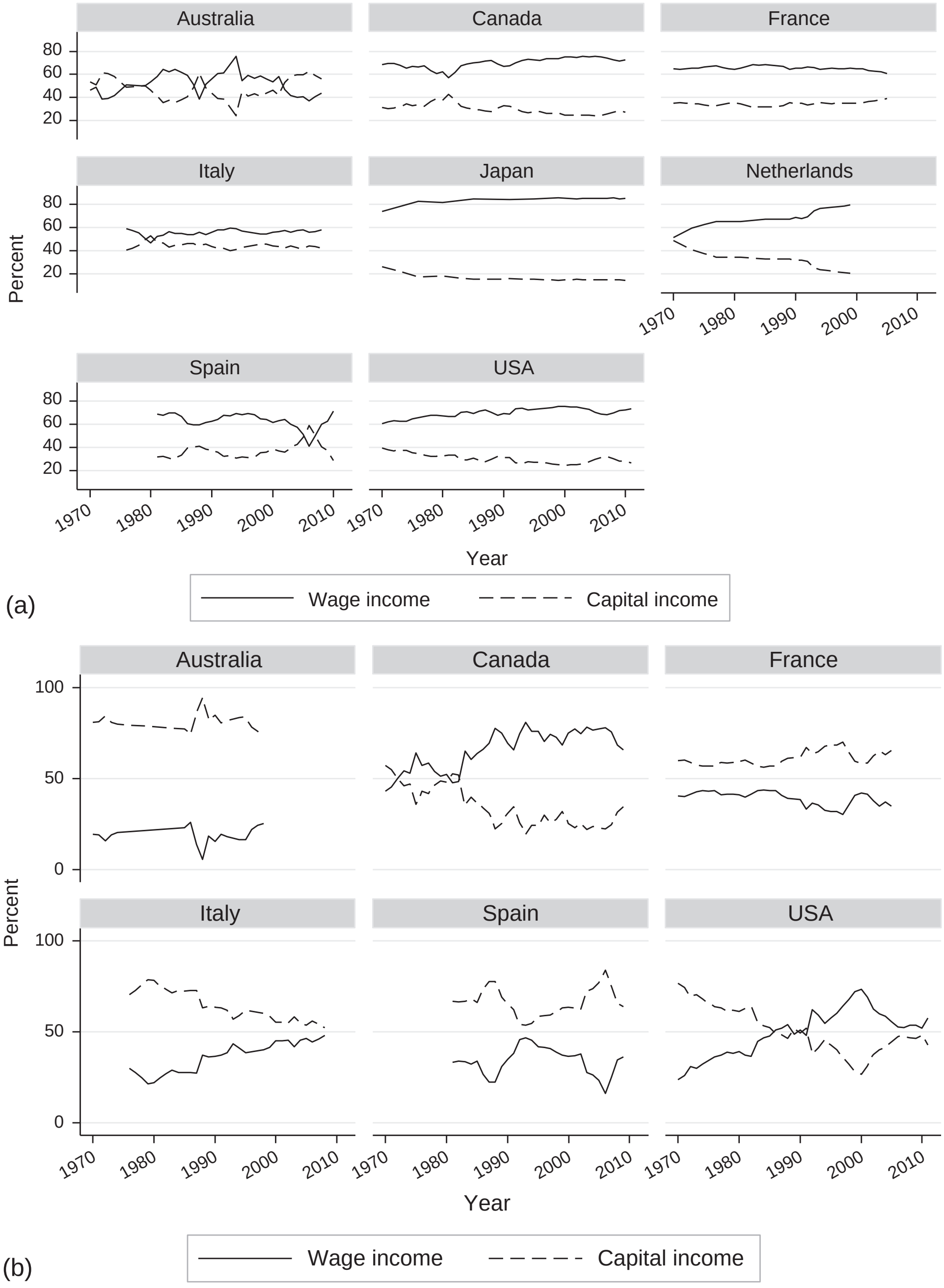


Figure 8.27 Labor and capital compositions of top incomes. (a) Income composition of the top 1% group. (b) Income composition of top 0.01% group. The graphs depict the incidence of different sources on total income accruing to the top 1%. In particular, the graphs show the share of capital income (including rental income, interest income, and dividends but excluding realized capital gains, where possible) and employment income (wages, salaries, bonuses, allowances, and pensions) for the eight countries where these calculations are possible (Australia, Canada, France, Japan, Italy, Netherlands, Spain, and the United States). Note that Australian and Spanish data include realized capital gains to the extent they are taxable. Moreover, Australian data can only be decomposed in wage and nonwage income ("capital" income includes realized capital gains as well as business income, self-employed income, profits from unincorporated businesses, and farm income). *Source: World Top Income Database, accessed September 2013. Elaboration by the authors.*

in wage shares. The clearest cases with increasing wage shares are the Netherlands after 1990 and Canada after 1980.

The evidence of an increase in top wages income is less clear cut for the top 0.01% shares (Figure 8.27b). Wage income visibly increased over time for Italy, Canada, and the United States (until around 2000, after which there was a marked reversal). The wage income share was relatively stable in Australia and France.

It is nonetheless important to bear in mind that definitions of income sources are not necessarily comparable across countries and that different tax systems within different countries may incentivize the reporting of a specific income source with greater fiscal convenience (see section 8.4.2.2.4 concerned with fiscal avoidance). It is therefore not always clear how one should interpret the documented percentage incidence of wage and capital incomes within top brackets. The results discussed within this section do not take these important issues into account.

8.4.2.6 Bridging the Gap Between Tax Statistics and Survey Data: Gini Versus Top Share

In the second section of this chapter we explored the dynamics of overall income inequality using a variety of summary statistics, including Gini coefficients, decile ratios, and others. These variables are usually constructed from household surveys. Some countries are, however, increasingly resorting to register data (such as the Scandinavian countries) or to a combination of both survey and register data (such as the United Kingdom and France since 2008) in an attempt to overcome standard household survey limitations.

Indeed, household surveys are not typically stratified by income, and, in part, as a result suffer limitations that are particularly pernicious for top income groups (measurement errors, nonresponse, or incomplete response); these surveys often adopt a top-coding methodology that by construction limits the information on the right tail of the income distribution.[72] These limitations frequently make it impossible to get robust quantitative evidence about the incomes of individuals at the very top of the distribution.[73]

[72] There are usually two levels of top-coding. The first level guarantees the secrecy of data, whereas the second (present even within the data exclusively available internally to statistical agencies) serves to avoid outliers having an influence on the aggregate statistics.

[73] This has been pointed out by Brewer et al. (2008) for the case of the United Kingdom, where the statistical office usually adjusts the measure of income accruing to the very rich individuals detected within the Family Resources Survey using the data provided within the Survey of Personal Income. In turn, the latter samples information on income from the tax administrative data of the HMRC and by design oversamples the information on rich individuals. Similarly, Moriguchi and Saez (2008) discuss the limitation of the National Survey of Family income and Expenditure (NSFIE), which collects a representative sample of about 10,000 households and, because it "contains few observations at the high end of income distribution, it is difficult to provide precise estimates for the top 0.5% income group and above using NSFIE data."

On the other hand, top income shares are constructed from tax administrative microdata or grouped tabulations and are particularly suitable to estimate the right tail of the income distribution. Nonetheless, they provide less compelling information about the bottom of the distribution.

Do these two different sources provide substitutable or complementary information? In other words, are top income shares to be combined with survey data to have a more complete picture of economic inequality within a country? Or, do top shares embed sufficient information to proxy the distribution of income as a whole? In what follows we discuss these important questions individually.

8.4.2.7 Are Top Income Shares Complementary to Household Survey Data?

Work by Burkhauser et al. (2012a) compared the evolution of the top income shares in the United States calculated in survey data (Current Population Survey (CPS) data from internal sources) with that provided by Piketty and Saez (2006) using administrative tax data (from the Internal Revenue Service).[74] Burkhauser et al. (2012a) suggest that CPS-based top shares track closely with the tax-based top shares up to the 99th percentile. Importantly, the comparison takes into account the same unit of reference (tax units) and a definition of income similar to that adopted in Piketty and Saez (2006). However, the U.S. top 1%, as estimated by Burkhauser et al. (2012a), does not track the top 1% obtained by Piketty and Saez (2006) with the same precision. This is even more evident once capital gains are included within the income definition,[75] as noted by Atkinson et al. (2011). Indeed, as shown before, capital gains are an important component of income at the top and could influence substantially both the level and the trend of income inequality. In addition, including capital gains arguably provides a more economically meaningful measure of income dispersion.[76]

Atkinson et al. (2011) also provide a tentative adjustment of official CPS Gini coefficients, taking into consideration the differentials in top 1% shares between survey-based and tax-based estimates (including capital gains). The result suggests that the official CPS data on Gini (household equalized gross income) fail to capture about half of the increase in overall inequality in the United States as measured by the adjusted Gini index.

[74] The internal CPS files contain better coverage of top incomes, specifically nontop-coded income levels, than the publicly available CPS.

[75] The CPS does not record capital gains. In addition, CPS data also exclude information on stock option gains, which are instead recorded within IRS taxation data.

[76] Moreover, the series including capital gains is less sensitive to changes in tax avoidance around TRA 1986. Indeed, as we discussed previously, the TRA 1986 mostly incentivized "a shift from corporate income toward individual business income." However, as noted by Atkinson et al. (2011), it is also true that "Before TRA 1986, small corporations retained earnings and profits accrued to shareholders as capital gains eventually realized and reported on individual tax returns. Therefore, income including capital gains does not display a discontinuity around TRA 1986 (1986 is artificially high due to high capital gains realizations before capital gains tax rates went up in 1987)."

These findings seem to indicate that taxation data are able to capture additional information that is not recorded within statistical surveys. Yet it is also important to stress here that the extent to which estimates based on survey data can be adjusted using tax statistics is not yet fully understood or investigated. Moreover, the required "adjustments" may well be different across countries. Such adjustments are increasingly implemented within the literature.

Atkinson (2007) provides the intuitive formal *approximate* relationship between the top share and the Gini coefficient, $G=(1-S)G^*+S$, where G represents the overall Gini coefficient, S is the top share, and G^* is the Gini coefficient for the rest of the population excluding the top individuals. However, the above-mentioned derivation requires the assumption that the top income group refers to an infinitesimal share of the population (say top 1%, top 0.1%, or top 0.01%). Alvaredo (2011) subsequently obtains the more general derivation valid for noninfinitesimal top groups as well: $G=(1-S)(1-P)G^*+S-P+G^{**}PS$, where P is the population share of the top group under investigation and G^{**} is the Gini relative to the distribution of income within the top group (G^{**} can be further simplified to $1/(2-\alpha)$, assuming that the right tail is Pareto distributed with coefficient α).

Under the presumption that the observed Gini coefficient (obtained from standard survey data) is a better representation of inequality within the bottom group (G^*), we can use the above results to obtain estimates of the adjusted Gini for the whole population (G). This could be considered to be the first approximate attempt to correct the overall measure of inequality using the available additional information about the top income share (for instance, this was illustrated for the case of Argentina by Alvaredo, 2011). From the discussion above one can already expect the actual value of the adjusted measure of Gini to depend on the choice of top shares to be used.

Assuming that the top percentile is excluded from the national household survey, we can illustrate the adjustment of the official Gini coefficient of gross equalized household income (including cash transfers) from the CPS data in the case of the United States. Using Atkinson's (2007) original formula illustrated above and the top 1% share from the WTID, including capital gains, the adjustment is worth 5 percentage points in 1970 and more than 10 percentage points in 2006 (Figure 8.28). The adjustment is approximately 1 percentage point lower if we use the top 1% excluding capital gains. Furthermore, using Alvaredo's (2011) more general formula, the adjustment is further reduced by one additional percentage point.

Such adjustments, however, depend on the strong assumption (not necessarily true) about the exact share of national income thought to be excluded from the household survey statistics. Such an assumption has to be carefully assessed before carrying out any corrections to the shares. Indeed, one could obtain adjusted Gini measures in a slightly more sophisticated way by estimating the top shares using both the survey and the

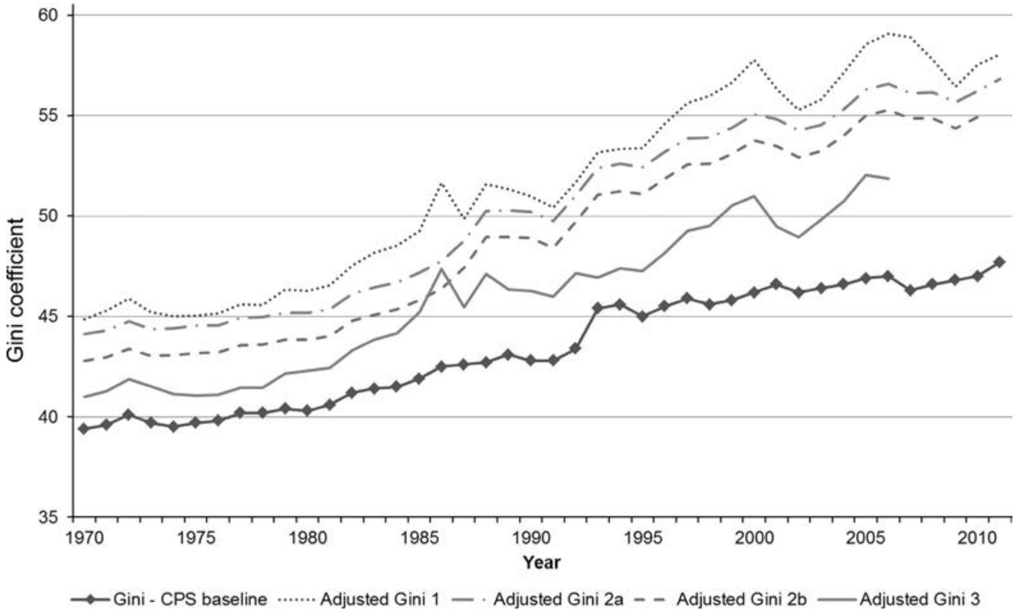


Figure 8.28 Adjusting the U.S. Gini coefficient using the top income shares. The baseline Gini coefficient represents the headline series of the U.S. Current Population Survey (CPS) based on household equivalized gross income. Top income shares estimated by Piketty and Saez (2003), Saez (2013), and Burkhauser et al. (2009) then are used to calculate adjusted measures of Gini coefficient. We carried out four different adjustments. Adjustment 1 assumes that the top 1% (including capital gains) is not captured at all within the household survey statistics, and we use the formula described by Atkinson (2007) to derive the "true" Gini coefficient: $G=(1-S)G^{*}+S$, where G represents the overall Gini coefficient, and S is the top share, and G^{*} is the Gini coefficient for the rest of the population excluding the top individuals. Adjustment 2a uses the same formula above but the top 1% excluding capital gains. Adjustment 2b makes use of the more general specification highlighted by Alvaredo (2011): $G=(1-S)(1-P)G^{*}+S-P+G^{**}PS$, where P is the population share of the top group under investigation and G^{**} is the Gini relative to the distribution of income within the top group (G^{**} can be further simplified to $1/(2-\alpha)$, assuming that the right tail is Pareto distributed with coefficient α). Finally, adjustment 3 assumes that the top 1% share is partially captured within the national survey. The difference D represents the estimates of top shares using taxation statistics (and including capital gains) from those using survey data; the following correction is then used to apply the following adjustment: $G=(1-D)G^{*}+D$. *Sources: Burkhauser et al. (2012), Atkinson et al. (2011), and calculations of the authors.*

taxation administrative data, adopting a homogeneous methodology (e.g., unit of analysis, income definition, control totals). The formulas mentioned above will then serve to adjust the available Gini from the survey data using the difference D of the estimates of top shares using different sources: $G=(1-D)G^{*}+D$. A similar adjustment was carried out in the case of the United States (discussed above) and illustrated by Atkinson et al. (2011).[77] This adjustment, using top income shares including capital gains, is represented in Figure 8.28 and represents a substantially smaller adjustment to the Gini coefficient than the ones discussed above (an approximate change of 1.5 percentage points in 1970 and 5 points in 2006 with respect to the actual baseline Gini coefficient based on gross income). This suggests that more than half of the increase in inequality from 1970 to 2006 is not captured by the inequality measure (based on household surveys), which excludes a sizeable part of the top 1% share of national income, as estimated with taxation data.[78]

Finally, one could go beyond the first approximation adjustments we discussed above by matching the individual information within surveys using administrative data with full coverage of the population. Not much research has been carried out yet at this stage, and this remains an important open issue that will attract the attention of economists and statisticians in the coming years.

To conclude, top income shares tend to be underestimated within household surveys (especially above the 99th percentile), and we have shown that taxation data can, in some cases, provide additional and complementary information that could not be otherwise recorded. Given the relentless increase in top income shares in many advanced and developing countries, it is possible that the official indicators of income inequality might substantially and increasingly underestimate the extent of the change in the actual dispersion of income distribution. However, data on reported income for taxation purposes are not without caveats, as we extensively discussed in this chapter, and caution is prudent when applying any kind of approximate correction to a Gini coefficient that is heavily dependent on arbitrary choices.

8.4.2.8 Changes in Top Income Shares as Proxies for the Overall Income Distribution

We discussed above how top income shares, as measured using tax statistics, may not be fully represented within household survey data. However, changes in top shares may still be informative about the dynamics of the income distribution as a whole, especially if much of the action is at the top, as suggested by the burgeoning literature on top incomes.

The analysis of data in this chapter highlighted how different measures of inequality generally result in similar impressions of how inequality has changed and which countries have the most unequal distributions of income, whether based on the entire distribution or only on top-income households. For instance, it is true that countries with larger top 1% shares also tend to have higher Gini coefficients, S80/S20 ratios, and P90/P10 ratios, although the correlation coefficient is substantially less than 1. The top 1% share (using

[77] However, the measure of the Gini used for the correction by Atkinson et al. (2011) and here is not computed from the survey using a methodology comparable with that of tax statistics. On the contrary, for illustrative purposes, they merely took the baseline Gini from the official publication of the CPS data.

[78] This was also recalled by Atkinson et al. (2011).

pretax income of tax units) and the Gini coefficient (using DHI) have a correlation coefficient of 0.65; using the S80/S20 ratio the correlation is also 0.65, and using the P90/P10 ratio, it is 0.72.[79]

Consistent with this evidence, an important study by Leigh (2007) found that the correlation between top shares and Gini coefficients[80] is not only strong in the cross section but also after controlling for country fixed effects and common time effects. This suggests that within-country changes of top income shares and Gini coefficients also are strongly correlated.

Given the importance of the work by Leigh (2007), and similar to what was done in Chapter 7, we extend here his analysis by making use of updated data on both top income shares and Gini coefficients made available since Leigh has published his work.[81] However, we further extend the work in several respects. First, we used two additional series of Gini coefficients. Specifically, we use the Gini series (related to equivalized DHI) assembled by Atkinson and Morelli (2012, 2014) within the *Chartbook of Economic Inequality*,[82] and we make use of series of Gini coefficients of gross/market income,[83] which is more directly comparable with the series of top income shares. Although Leigh's analysis stretches back to the early years of twentieth century, we focus here on the post-1970 period only.

Second, and most important, we acknowledge here that Leigh's (2007) original specification treats top income shares as the dependent variable. To the extent that we need to analyze the informative content of top shares for the overall income distribution measures, this is not necessarily the preferred approach.[84] Thus we reverse the order of the regression variables by regressing the log of Gini on the log of top shares to obtain more direct information about the elasticity of the Gini coefficient as it relates to changes in top share. This is shown in Table 8.7, where the elasticity of Gini to changes in top

[79] The simple pairwise correlations between the different inequality measures were calculated using the figures from Table 8.2 and Figure 8.24.

[80] The study by Leigh (2007) made use of Gini coefficients of DHI from the LIS and from the World Income Inequality Database (WIID).

[81] For evidence of the relationship between Gini coefficients and top income shares for developing countries we direct the reader to Chapter 9.

[82] The *Chartbook of Economic Inequality* covers the more than 100-year period since 1900 for 25 different countries, accounting for more than a third of the world population. The database collects information on five different annual measures of "economic inequality," among which is the Gini index on equivalized DHI (the remaining measures are top income shares, income- or consumption-based poverty measures, earnings dispersion measures, and top wealth shares). The underlying figures are available for download at: www.chartbookofeconomicinequality.com.

[83] These series, with few exceptions, are mostly retrieved from the OECD inequality database, accessed on October 1, 2013.

[84] The regression order of two variables is irrelevant only if the variables under investigation are standardized. Furthermore, the choice of the order of regression is essentially a scientific one and reflects the nature of the theory underlying the empirical specification.

Table 8.7 Assessing the elasticity of Gini coefficients to changes in top 1% shares

	Pooled OLS regression		OLS regression with country fixed effects		OLS regression including country and time effects	
Regressing log (Gini) on log (top 1%): data from 1970 to 2011						
Gini (disposable income)—LIS	0.3 (0.03)⋆		0.2 (0.03)⋆		0.2 (0.1)⋆	
Gini (disposable income)—WIID		0.2 (0.04)⋆		0.2 (0.03)⋆		0.2 (0.1)⋆
R^2	0.50	0.13	0.91	0.65	0.93	0.80
N	103	373	103	373	103	373
Gini (disposable income)—*Chartbook*	0.4 (0.02)⋆		0.3 (0.01)⋆		0.3 (0.03)⋆	
Gini (gross income)—OECD		0.1 (0.02)⋆		0.2 (0.01)⋆		0.01 (0.03)
R^2	0.45	0.05	0.93	0.80	0.93	0.91
N	343	245	343	245	343	245

Note: A significance level of 1% is indicated by ⋆.

shares is estimated to be in the range of 10–40%, with strong statistical significance if we consider the whole post-1970 period.

As a third step, we also estimate the elasticity of Gini coefficients to changes in the top 1% across three different subperiods: 1970–1985, 1986–2000, and 2001–2012. This allows us to study the evolution of the elasticity of the Gini to change in the top 1% over time.[85] The findings show that the relationship between changes in top shares and changes in the Gini coefficient tends to disappear during the latest period (Table 8.8).[86] This seems consistent with the facts observed within previous sections, where top shares show no sign of having "peaked," whereas Gini coefficients have increased at a slower pace in many countries since the 1980s or the 1990s. One reason, as discussed before, may be that household income surveys poorly measure the top share.[87]

[85] Different from Leigh (2007), we also focus exclusively on the top 1% and not on the top 10%. Moreover, we check the relationship between top 1% shares and Gini coefficients only, excluding other types of inequality measures, such as Atkinson's indices and interfractile ratios.

[86] It is worth noting that the relationship between the top 1% and Gini is generally more robust over time using data from the *Chartbook of Economic Inequality*. For simplicity, Table 8.8 exclusively represents the evidence about Gini based on gross income from the OECD and Gini based on net income from LIS data.

[87] For instance, Kenworthy and Smeeding (2013) show that, in the United States, once the top centile is removed from survey data, income inequality has increased much more modestly since the mid-1990s. These findings suggest that, viewed cross-sectionally, the increase in the top end has driven much of the distribution in the United States.

Table 8.8 Assessing the elasticity of Gini coefficients to changes in top 1% shares over time

	Period 1: 1970–1985	**Period 2: 1986–2000**	**Period 3: 2001–2012**
Gini (disposable income)—LIS	0.2 (0.07)⋆	0.2 (0.05)⋆	−0.1 (0.2)
R^2	0.95	0.60	0.93
N	22	60	60
Gini (gross income) —OECD	0.1 (0.04)⋆	0.2 (0.04)⋆	0.03 (0.05)
R^2	0.85	0.98	0.90
N	61	98	86

Note: A significance level of 1% is indicated by ⋆. Regressing log(Gini) on log(top1) using fixed effects regression with robust SEs.

Finally, we also replicate[88] here the original specification (regressing log of top shares on the log of Gini) by Leigh (2007). Results are represented in Table 8.9, where the original results by Leigh (panel A) are compared with those making use of more up-to-date data (panel B) as well as with the results based on two different series of Gini coefficients (panel C).

It is worth noting that the use of up-to-date and adjusted inequality series, together with the restriction to the post-1970 period, does not seem to affect the validity of Leigh's (2007) findings (panel B). Similarly, the use of the two additional series of Gini coefficients (panel C) substantially confirms the Leigh's findings.[89] The latter result is relevant because Gini coefficients based on pretax and pretransfer income are more appropriate data series to compare with top income shares (based on gross income).[90]

To summarize, the relationship between changes in the top shares and Gini coefficients documented by Leigh (2007) remains strong and robust to the controls for updated information, restricted period sample, and different Gini indicators, including that based on pretax and pretransfer income. Hence, changes in top income shares remain, on average, a good proxy for overall income distribution despite the misrepresentation of top income brackets within the statistical survey data documented above. However, there is evidence suggesting that the relationship between Gini and top shares became weaker during the first decade of the twenty-first century, suggesting that household surveys may not entirely capture the dynamics of income at the top. This suggests that greater

[88] We thank Andrew Leigh for kindly providing the original code, which eased the replication of the findings.

[89] The strong association between changes in top 1% shares and Gini coefficients is only weakened once the analysis is restricted to the use the Gini coefficients based on the definition of gross income, and we use the time and fixed effects specification. Indeed, the elasticity coefficient based on the two-way fixed effects regressions between the top 1% and gross Gini is no longer statistically different from zero.

[90] Results are broadly consistent using the top 10% share, although results are not tabulated. The findings are available upon request.

Table 8.9 Assessing the association between the top 1% and Gini coefficients, replicating the findings by Leigh (2007)

	Pooled OLS regression		OLS regression with country fixed effects		OLS regression including country and time effects	
Panel A: Original findings by Leigh (2007)[a]						
Gini (disposable income)—LIS	1.45 (0.203)*		1.19 (0.298)*		0.797 (0.62)	
Gini (disposable income)—WIID		0.799 (0.086)*		0.693 (0.1)*		0.422 (0.07)*
R^2	0.44	0.29	0.83	0.67	0.96	0.89
N	63	300	63	300	63	300
Panel B: Original specification by Leigh (2007) re-run with up-to-date observations[b]						
Gini (disposable income)—LIS	1.5 (0.1)*		2.1 (0.3)*		0.7 (0.3)*	
Gini (disposable income)—WIID		0.6 (0.1)*		0.4 (0.1)*		0.2 (0.1)*
R^2	0.50	0.13	0.77	0.66	0.93	0.85
N	103	373	103	373	103	373
Panel C: Original specification by Leigh (2007) re-run using different Gini series[b]						
Gini (disposable income)—*Chartbook*	1.0 (0.1)*		1.9 (0.1)*		1.0 (0.04)*	
Gini (gross income)—OECD		0.7 (0.2)*		2.3 (0.2)*		0.1 (0.3)
R^2	0.45	0.02	0.82	0.81	0.90	0.93
N	343	245	343	245	343	245

Note: A significance level of 1% is indicated by *.
[a]All observations available from 1886 to 2004—regressing log (top 1%) on log (Gini).
[b]Data from 1970 to 2011—regressing log (top 1%) on log (Gini).

prudence is called for when extrapolating the validity of any results based on the analysis of top income shares directly to the overall income distribution.[91]

8.5. SUMMARY AND CONCLUSIONS

This chapter focused on a wealth of new inequality data that has grown in detail, form, and importance since 2000. In addition to LIS, which was the bedrock of work by

[91] This conclusion seems to differ from what was discussed in Chapter 7, where the focus on the long-run relationship between top shares and Gini coefficients does not allow the weakening of such a relationship to be identified for the first decade of the twenty-first century.

Gottschalk and Smeeding (2000), OECD, EU-SILC, and a series of country trend data, maintained in part by Atkinson and Morelli (2012, 2014) and Brandolini and Smeeding (2008, 2009), have made a tremendous difference in what we know about levels and trends in inequality and poverty over the past 30 years. Importantly, a whole new set of WTID data has proliferated and offers long-term trends in inequality for tax units at the top of the distribution. All of these new data come with new complications and caveats, which were discussed in the previous sections. Despite these caveats, these data do allow us to make a few overall summarizing observations about levels and trends in poverty and inequality over the past 30 years.

The modest conclusions we draw here include the following:

1. Of 28 rich and MIC nations in the late 2000s, 17 nations successfully reached single-digit poverty rates (where between 5% and 10% of the country's population are poor by the half-median relative poverty measure). But the range of poverty rates in rich nations alone varies by a factor of almost 4 and, when adding MICs, by a factor of 5. Hence one experiences a wide range of relative poverty rates in these nations. Our trend data suggest that progress against poverty was uneven and rare in rich nations over the past 20–30 years. Other than Mexico, relative poverty rates did not consistently decrease over the past 15–20 years in any of the nations we examine here.
2. We conclude that while there was little progress in reducing relative poverty in almost all the rich countries examined here over the past two or three decades (up to 2010), real living standards for the poor have changed over this same period. Anchored poverty is an increasingly useful concept to establish how upward and downward changes in real median incomes affect poverty differently from a solely relative measure. Anchored poverty decreased in almost all rich nations from the 1990s to 2007 because of rising living standards in most of the rich world up to that point. However, since the GR, increases in anchored poverty up to 2010 reduced some of the progress in real living standards that low-income households experienced over the preceding 15 years.
3. Inequality increased (almost) everywhere over the 1970–2010 period, with some flattening during the GR, although the longer-term rising trend continued. Small changes year to year may produce strong trends over a 20–30-year period. Long-term increases in the Gini coefficients, P90/P10 ratios, and S80/S20 ratios are evident for DHI calculated using household surveys and with top income shares calculated with tax data.
4. The cyclicality of some measures of inequality—particularly for top income shares—is demonstrated clearly in the trends calculated with the WTID. Recessions have depressed the incomes of the rich especially, but these incomes bounced back even stronger after the recessions of the last decades of the twentieth century. Preliminary evidence suggests the same pattern will likely hold for the GR.

5. The 1950–1980 period stands out as the "golden age" for labor and decreasing or stable inequality in the rich western nations. Several additional nations now show a U-shaped pattern of inequality, with inequality increasing even more since the last look at this phenomenon 14 years ago (Gottschalk and Smeeding, 1997, 2000; OECD, 2011). The longer time series of the WTID shows an even stronger U shape in inequality trends in these data.
6. Cross-national inequality rankings in the most recent data largely look similar to how they appeared 15 or even 30 years ago. The English-speaking countries (led by the United States and the United Kingdom) are the most unequal, by most measures, and the Nordic countries are the least unequal. There have been some important changes to note as well. New data allow us to add Israel and South Africa to the list of the most unequal. Also, the distance between the most and least unequal among rich countries has diminished as inequality growth surged in some of the least unequal.
7. Increasingly, one has to examine capital income as well as earned income. Increasing income from capital is more concentrated at the top of the distribution, as seen in the WTID in many nations since the 1990s and through the GR.
8. Broad-based distribution measures increased most in the 1970s, 1980s, or 1990s (depending on the country) but rose less, and were sometimes stable, in the 2000s. Using top income shares, however, inequality seems to still be rising and shows no sign of having "peaked." How long this pattern can continue is an open question.
9. The relentless increase in top income shares poses new challenges to the informative content of different indicators of income inequality. On the one hand, intrinsic limitations of existing household surveys do not capture the entirety of income accruing to the top income brackets. This suggests that conventional measures such as the Gini coefficient may be increasingly missing the actual extent of the change in income inequality. On the other hand, there is evidence suggesting that the relationship between Gini and top shares became weaker over the past decade, pointing to greater prudence in extrapolating any results based on the analysis of top income shares directly to the overall income distribution.

The future research agenda for empirical studies of inequality and poverty is quite rich and may provide the answers to many questions that are not clear at this point. Additional research on the relationship of inequality and economic growth, as well as who receives the growth dividends, is called for. In a rich and aging world, how will changes in the age distribution affect inequality? In addition, and perhaps most important, the increasing availability and usefulness of data from MICs will provide us with comparisons to the living standards in these and poorer countries. Suffice it to say that with inequality increasing in most rich nations and with increased coverage of the top 1% of income earners, and of MICs, we still have much to learn about inequality, its sources, its origins, and its effects on social and economic outcomes. It is indeed time to bring inequality back into the fold of mainstream economics, as Atkinson (1997) suggested.

ACKNOWLEDGMENTS

The authors thank their organizations for support of this work. In particular, the authors especially thank David Chancellor and Dawn Duren of IRP for their assistance with graphs and manuscript preparation and Peter Frase for research assistance. The authors also thank Tony Atkinson, Francois Bourguignon, Michael Forster, Andrew Leigh, Max Roser, Stephen Jenkins, and the editors for helpful comments on an earlier draft and Andrea Brandolini for providing helpful inequality trend data. The authors, not their organizations, are responsible for all opinions in this chapter, as well as all errors of omission and commission.

REFERENCES

Aaberge, R., Atkinson, A.B., 2008. Top incomes in Norway, Statistics Norway Discussion Paper Series n.552.

Aaron, H., 1985. Comments on methods of measuring noncash benefits. In: Reported in Noncash Benefits: An Evaluation of the Census Bureau's Measurement Conference, PEMD-86-8BR. U.S. General Accounting Office, Washington, DC, April.

Abel-Smith, B., Townsend, P., 1965. The Poor and the Poorest. Bell, London.

Allison, P., 1978. Measures of inequality. Am. Sociol. Rev. 43 (6), 865–880.

Alvaredo, F., 2010. The Rich in Argentina over the Twentieth Century, 1932–2004. In: Atkinson, A.B., Piketty, T. (Eds.), Top Incomes: A Global Perspective. Oxford University Press, Oxford and New York.

Alvaredo, F., 2011. A note on the relationship between top income shares and the Gini coefficient. Econ. Lett. 110 (3), 274–277.

Alvaredo, F., Atkinson, A.B., 2010. Colonial Rule, Apartheid and Natural Resources: Top Incomes in South Africa, 1903-2007. CEPR Discussion Paper No. DP8155.

Alvaredo, F., Saez, E., 2009. Income and wealth concentration in Spain from a historical and fiscal perspective. J. Eur. Econ. Assoc. 7 (5), 1140–1167.

Alvaredo, F., Atkinson, A.B., Piketty, T., Saez, E., 2012. The World Top Incomes Database. Accessed in September 2013, http://topincomes.g-mond.parisschoolofeconomics.eu/.

Alvaredo, F., Atkinson, A.B., Piketty, T., Saez, E., 2013. The top 1 percent in international and historical perspective. J. Econ. Perspect. 27 (3), 3–20.

Atkinson, A.B., 1970. On the measurement of inequality. J. Econ. Theory 2, 244–263.

Atkinson, A.B., 1983. The Economics of Inequality, second ed. Clarendon Press, Oxford.

Atkinson, A.B., 1997. Bringing income distribution in from the cold. The Economic Journal, No. 107, March, pp. 297321.

Atkinson, A.B., 2005. Top incomes in the UK over the 20th century. J. R. Stat. Soc. A. Stat. Soc. 168 (2), 325–343.

Atkinson, A.B., 2007. Measuring top incomes: methodological issues. In: Atkinson, A.B., Piketty, T. (Eds.), Top Incomes over the Twentieth Century: A Contrast Between Continental European and English-Speaking Countries. Oxford University Press, Oxford, pp. 18–42.

Atkinson, A.B., Brandolini, A., 2001. Promises and pitfalls in the use of secondary data-sets: income inequality in OECD countries as a case study. J. Econ. Lit. 39, 771–800.

Atkinson, A.B., Harrison, A.J., 1978. The Distribution of Personal Wealth in Britain. Cambridge University Press, Cambridge, MA.

Atkinson, A.B., Leigh, A., 2005. The distribution of top incomes in New Zealand, Australian National University CEPR Discussion Paper 503.

Atkinson, A.B., Leigh, A., 2007. The distribution of top incomes in Australia. Econ. Rec. 83 (262), 247–261.

Atkinson, A.B., Marlier, E., 2010. Income and Living Conditions in Europe. Eurostat at, http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-31-10-555/EN/KS-31-10-555-EN.PDF.

Atkinson, A.B., Morelli, S., 2012. Chartbook of economic inequality: 25 countries 1911–2010. INET Research Note #15, Institute for New Economic Thinking, New York October.

Atkinson, A.B., Morelli, S., 2014. Chartbook of Economic Inequality. www.chartbookofeconomicinequality.com.
Atkinson, A.B., Piketty, T., 2007. Top Incomes over the 20th Century: A Contrast Between Continental European and English-Speaking Countries. Oxford University Press, Oxford.
Atkinson, A.B., Piketty, T., Saez, E., 2009. Top Incomes in the Long Run of History. NBER Working Paper No. 15408 Issued in October 2009.
Atkinson, A.B., Piketty, T., 2010. Top Incomes: A Global Perspective. Oxford University Press, Oxford and New York.
Atkinson, A.B., Salverda, W., 2005. Top incomes in the Netherlands and the United Kingdom over the 20th century. J. Eur. Econ. Assoc. 3 (4), 883–913.
Atkinson, A.B., Rainwater, L., Smeeding, T.M., 1995. Income Distribution in OECD Countries: The Evidence from the Luxembourg Income Study (LIS). Organization for Economic Cooperation and Development, Paris.
Atkinson, A.B., Cantillon, B., Marlier, E., Nolan, B., 2002. Social Indicators: The EU and Social Inclusion. Oxford University Press, Oxford.
Atkinson, A.B., Piketty, T., Saez, E., 2010a. Top incomes in the long run of history. In: Atkinson, A.B., Piketty, T. (Eds.), Top Incomes: A Global Perspective. Oxford University Press, Oxford and New York, pp. 664–759.
Atkinson, A.B., Marlier, E., Montaigne, F., Reinstadler, A., 2010b. Income poverty and income inequality. In: Atkinson, A.B., Marlier, E. (Eds.), Income and Living Conditions in Europe. Publications Office of the European Union, Luxembourg.
Atkinson, A.B., Piketty, T., Saez, E., 2011. Top incomes in the long run of history. J. Econ. Lit. 49 (1), 3–71.
Banerjee, A., Piketty, T., 2005. Top Indian incomes, 1922–2000. World Bank Econ. Rev. 19 (1), 1–20.
Björklund, A., 1993. A comparison between actual distributions of annual and lifetime income: Sweden 1951–89. Rev. Income Wealth 39, 377–386.
Björklund, A., Freeman, R., 1997. Generating equality and eliminating poverty—the Swedish way. In: Freeman, R.B., Topel, R., Swedenborg, B. (Eds.), The Welfare State in Transition: Reforming the Swedish Model. University of Chicago Press, Chicago.
Björklund, A., Palme, M., 2002. Income redistribution within the life cycle versus between individuals: empirical evidence using Swedish panel data. In: Cohen, D., Piketty, T., Saint-Paul, G. (Eds.), The Economics of Rising Inequalities. Oxford University Press, Oxford, pp. 205–223.
Blank, R., 2008. How to improve poverty measurement in the United States. J. Policy Anal. Manage. 27 (2), 233–254.
Böheim, R., Jenkins, S.P., 2006. A comparison of current and annual measures of income in the British Household Panel Survey. J. Off. Stat. 22, 733–758.
Boltvinik, J., 2000. Poverty Measurement and Trends. United Nations Development Programme, SEPED Series on Poverty Reduction. United Nations, New York.www.undp.org/poverty/publications/pov_red.
Booth, C., 1903. Life and Labour of the People of London. Second Series, Industry, vol. 5. Macmillan and Co., London
Brandolini, A., Smeeding, T.M., 2008. Inequality patterns in western-type democracies: cross-country differences and time changes. In: Beramendi, P., Anderson, C.J. (Eds.), Democracy, Inequality and Representation. Russell Sage Foundation, New York.
Brandolini, A., Smeeding, T.M., 2009. Income inequality in richer and OECD countries. In: Salverda, W., Nolan, B., Smeeding, T.M. (Eds.), Oxford Handbook of Economic Inequality. Oxford University Press, Oxford, pp. 71–100.
Brandolini, A., Vecchi, G., 2011. The well-being of Italians: a comparative historical approach, Bank of Italy Quaderni di Storia Economica No. 19, October.
Brandolini, A., Magri, S., Smeeding, T.M., 2010. Asset-based measurement of poverty. J. Policy Anal. Manage. 29 (2), 267–284.
Brewer, M., Muriel, A., Phillips, D., Sibieta, D., 2008. Poverty and Inequality in the UK: 2008. Institute for Fiscal Studies, London.

Buhmann, B., Rainwater, L., Schmaus, G., Smeeding, T.M., 1988. Equivalence scales, well-being, inequality, and poverty: sensitivity estimates across ten countries using the Luxembourg Income Study (LIS) Database. Rev. Income Wealth 34, 115–142.

Burkhauser, R.V., Feng, S., Jenkins, S.P., 2009. Using the P90/P10 index to measure US inequality trends with current population survey data: a view from inside the Census Bureau vaults. Rev. Income Wealth 55 (1), 166–185.

Burkhauser, R.V., Feng, S., Jenkins, S.P., Larrimore, J., 2012a. Recent trends in top income shares in the USA: reconciling estimates from March CPS and IRS tax return data. Rev. Econ. Stat. 94 (2), 371–388.

Burkhauser, R.V., Larrimore, J., Simon, K., 2012b. A second opinion on the economic health of the American middle class and why it matters in gauging the impact of government policy. Natl. Tax J. 65, 7–32.

Burkhauser, R.V., Hahn, M.H., Wilkins, R., 2013. Measuring top incomes using tax record data: a cautionary tale from Australia, Melbourne Institute Working Paper Series, Working Paper No. 24/13.

Caminada, K., Goudswaard, K., Wang, C., 2012. Disentangling income inequality and the redistributive effect of taxes and transfers in 20 LIS countries over time, Luxembourg Income Study Working Paper No. 581, September.

Canberra Group, 2001. Final Report and Recommendations of the Canberra Expert Group on Household Income Statistics, Ottawa, Canada.

Canberra Group, 2011. Canberra Group Handbook on Household Income Statistics, second ed. Canberra Group, Geneva. athttp://www.unece.org/fileadmin/DAM/stats/groups/cgh/Canbera_Handbook_2011_WEB.pdf.

Chen, S., Ravallion, M., 2012. Absolute poverty measures for the developing world. In: Measuring the Real Size of the World Economy. World Bank, Washington, DC (Chapter 20).

Clark, A., d'Ambrosio, C., 2014. Attitudes to income inequality: experimental and survey evidence. In: Atkinson, A.B., Bourguignon, F. (Eds.), Handbook of Income Distribution. In: vol. 2A. Elsevier B.V. (Chapter 13), December.

Cobham, A., Sumner, A., 2013. Is it all about the tails? The Palma measure of income inequality, Center for Global Development, Working Paper 343, September 2013.

Coleman, J., 1988. Social capital in the creation of human capital. Am. J. Sociol. 94, S95–S120.

Coleman, J., Rainwater, L., 1978. Social Standing in America. Basic Books, New York.

Coulter, F.A.E., Cowell, F.A., Jenkins, S.P., 1992. Equivalence scale relativities and the extent of inequality and poverty. Econ. J. 102, 1067–1082.

Cowell, F.A., 2000. Measurement of inequality. In: Atkinson, A.B., Bourguignon, F. (Eds.), Handbook of Income Distribution, vol. 1. North Holland, Amsterdam, pp. 87–166.

Cowell, F.A., Victoria-Feser, M.P., 1996. Robustness properties of inequality measures. Econometrica 64, 77–101.

Danziger, S., Taussig, M.K., 1979. The income unit and the anatomy of income distribution. Rev. Income Wealth 25, 365–375.

Deaton, A., Grosh, M., 2000. Consumption. In: Grosh, M., Glewwe, P. (Eds.), Designing Household Survey Questionnaires for Developing Countries. Lessons from 15 Years of the Living Standards Measurement Study. In: vol. 1. World Bank, Washington, DC, pp. 91–133.

Deding, M.C., Dall Schmidt, T., 2002. Differences in income inequality across Europe—market driven or . . .? European Panel Analysis Group Working Paper 2002-37, December.

DeNavas-Walt, C., Proctor, B.D., Smith, J.C., 2012. Income, Poverty, and Health Insurance: 2011. Current Population Reports, P60-243, U.S. Census Bureau, at http://www.census.gov/prod/2012pubs/p60-243.pdf.

Department for Work and Pensions (UK), 2012. Households below average income. Available at, http://research.dwp.gov.uk/asd/index.php?page=hbai.

Dhongde, S., 2013. Measuring multidimensional poverty in the U.S. Presentation at Institute for Research on Poverty, University of Wisconsin, Madison, March 20.

Ebert, U., 1997. Social welfare when needs differ: an axiomatic approach. Economica 64, 233–244.

Eurostat, 1998. Recommendations of the Task Force on Statistics on Social Exclusion and Poverty. European Statistical Office, Luxembourg, October.

Eurostat, 2005. Income Poverty and Social Exclusion in the EU25, Statistics in Focus 03/2005. Official Publications of the European Communities, Luxembourg.
Eurostat, 2011. Social protection and social exclusion—the social dimension of the Europe 2020 strategy. Available at http://ec.europa.eu/social/main.jsp?catId=750&langId=en&pubId=5976&type=2&furtherPubs=yes.
Faik, J., 2012. Income inequality and poverty in front of and during the economic crisis—an empirical investigation for Germany, 2002–2010, SOEP Paper #450, DIW Berlin.
Ferreira de Souza, P.H.G., 2012. Poverty, inequality, and social policies in Brazil, 1995–2009, Working Paper No. 87, International Policy Centre for Inclusive Growth United Nations Development Programme, February.
Fisher, J., Johnson, D., Smeeding, T.M., 2012. Inequality of income and consumption: measuring the trends in inequality from 1985–2010 for the same individuals. In: Paper prepared for the 32nd General Conference of the International Association for Research in Income and Wealth, August.
Förster, M., 1993. Comparing poverty in 13 OECD countries: traditional and synthetic approaches. Studies in Social Policy 10. OECD, Paris, October.
Foster, J., Greer, J., Thorbecke, E., 1984. A class of decomposable poverty measures. Econometrica 3 (52), 761–766.
Frankel, S.H., Herzfeld, H., 1943. European income distribution in the Union of South Africa and the effect thereon of income taxation. S. AFR. J. Econ. 11 (2), 121–136.
Frick, J.R., Grabka, M.M., 2003. Imputed rent and income inequality: a decomposition analysis for Great Britain, West Germany, and the U.S. Rev. Income Wealth 49, 513–537.
Friedman, M., 1957. A Theory of the Consumption Function. Princeton University Press, Princeton, NJ.
Fuchs, V.R., 1967. Redefining poverty and redistributing income. Public Interest 8, 88–95.
Fuchs-Schündeln, N., Schündeln, M., 2009. Who stays, who goes, who returns? East-West migration within Germany since reunification. Econ. Transit. 17 (4), 703–738.
Garfinkel, I., Rainwater, L., Smeeding, T.M., 2006. A reexamination of welfare state and inequality in rich nations: how in-kind transfers and indirect taxes change the story. J. Policy Anal. Manage. 25, 855–919.
Garfinkel, I., Rainwater, L., Smeeding, T.M., 2010. Wealth and Welfare States: Is America a Laggard or Leader? Oxford University Press, Oxford.
Gilbert, G., 2008. Rich and Poor in America: A Reference Handbook. ABC-CLIO Press, New York.
Goebel, J., Gornig, M., Haubermann, H., 2010. Income polarization in Germany is rising. Weekly Rep. 6 (26)DIW Berlin.
Gordon, R.H., Slemrod, J.B., 2000. Are 'real' responses to taxes simply income shifting between corporate and personal tax bases? In: Slemrod, J.B. (Ed.), Does Atlas Shrug? The Economic Consequences of Taxing the Rich. Russell Sage Foundation, New York, pp. 240–280 Cambridge and London: Harvard University Press.
Gornick, J.C., Jäntti, M., 2013. Inequality and the Status of the Middle Class. Cambridge University Press, Cambridge, MA.
Gornick, J.C., Sierminska, E., Smeeding, T.M., 2009. The income and wealth packages of older women in cross-national perspective. J. Gerontol. 64B (3), 402–414.
Gottschalk, P., Smeeding, T.M., 1997. Cross-national comparisons of earnings and income inequality. J. Econ. Lit. 35, 633–687.
Gottschalk, P., Smeeding, T.M., 2000. Empirical evidence on income inequality in industrialized countries. In: Atkinson, A.B., Bourguignon, F. (Eds.), Handbook of Income Distribution, vol. 1. North-Holland, Amsterdam, pp. 261–308.
Grabka, M., Kuhn, U., 2012. The evolution of income inequality in Germany and Switzerland since the turn of the millennium, SOEP Paper #464, DIW Berlin.
Groedhart, T., Halberstadt, V., Kapteyn, A., Van Praag, B., 1977. The poverty line: concept and measurement. J. Hum. Resour. 12, 503–520.
Hagenaars, A.J.M., van Praag, B.M.S., 1985. A synthesis of poverty line definitions. Rev. Income Wealth 31 (2), 139–154.
Hagenaars, A.J.M., de Vos, K., Zaidi, M.A., 1994. Poverty Statistics in the Late 1980s: Research Based on Micro-Data. Eurostat, Luxembourg.

Haig, R.M., 1921. The concept of income: economic and legal aspects. In: Haig, R.M. (Ed.), The Federal Income Tax. Columbia University Press, New York, pp. 1–28.
Haveman, R., 2009. What does it mean to be poor in a rich society? In: Cancian, M., Danziger, S. (Eds.), Changing Poverty, Changing Policies. Russell Sage Foundation, New York.
Heshmati, A., 2004. Inequalities and their measurement. IZA Discussion Paper No. 1219, July.
Iacovou, M., Kaminska, O., Levy, H., 2012. Using EU-SILC data for cross-national analysis: strengths, problems, and recommendations, ISER Working Paper No. 2012-03.
Immervoll, H., Richardson, L., 2011. Redistribution policy and inequality reduction in OECD countries: what has changed in two decades? Luxembourg Income Study Working Paper No. 571, October.
Internal Revenue Service, U.S. Department of the Treasury, 1996. Federal Tax Compliance Research: Individual Income Tax Gap Estimates for 1985, 1988, and 1992, Pub. 1415 (rev. 4–96), April, Washington, DC.
Internal Revenue Service, U.S. Department of the Treasury, 2006. Updated Estimates of the TY 2001 Individual Income Tax Underreporting Gap, Overview. Office of Research, Analysis, and Statistics, Washington, DC, February 22.
International Monetary Fund, 2013. World Economic Outlook, April 2013 ed. International Monetary Fund, Washington, DC. Data available at http://www.imf.org/external/pubs/ft/weo/2013/01/.
Iwamoto, Y., Fujishima, Y., Akiyama, N., 1995. Rishi haito kazei no hyoka to kadai (Evaluating interest and dividends taxation). Finansharu Rebyu 35, 27–50.
Jäntti, M., Danziger, S., 2000. Income poverty in advanced countries. In: Atkinson, A.B., Bourguignon, F. (Eds.), Handbook of Income Distribution, vol. 1. North-Holland, Amsterdam, pp. 309–378.
Jäntti, M., Riihelä, M., Sullström, R., Tuomala, M., 2010. Trends in top income shares in Finland. In: Atkinson, A.B., Piketty, T. (Eds.), Top Incomes: A Global Perspective. Oxford University Press, Oxford and New York.
Johnson, D.S., Smeeding, T.M., 2012. A consumer's guide to interpreting various U.S. poverty measures. In: Fast Focus 14-2012. Institute for Research on Poverty, Madison, WI Available at, http://www.irp.wisc.edu/publications/fastfocus.htm.
Johnson, D.S., Smeeding, T.M., 2013. Inequality measurement, International Encyclopedia of the Social and Behavioral Sciences, second ed. Elsevier B.V.
Johnson, D.S., Smeeding, T.M., Torrey, B.B., 2005. Economic inequality through the prisms of income and consumption. Mon. Labor Rev. 128 (4), 11–24.
Kenworthy, L., 1998. Do social-welfare policies reduce poverty? A cross-national assessment. Luxembourg Income Study Working Paper #188Center for Policy Research, Syracuse University, Syracuse, NY. September, http://lissy.ceps.lu/wpapersd.htm.
Kenworthy, L., Smeeding, T.M., 2013. The United States: high and rapidly rising inequality. U.S, country chapter prepared for 2013 GINI Project, January.
Kuznets, S., 1953. Shares of Upper Income Groups in Income and Savings. National Bureau of Economic Research, New York.
Lampman, R., 1964. The problem of poverty in America. In: Economic Report of the President, 1964. U.S. Government Printing Office, Washington, DC (Chapter 2).
Leigh, A., 2007. How closely do top income shares track other measures of inequality? Econ. J. 117 (524), 619–633.
Leigh, A., 2009. Top incomes. In: Salverda, W., Nolan, B., Smeeding, T.M. (Eds.), The Oxford Handbook of Economic Inequality. Oxford University Press, Oxford and New York.
Leigh, A., van der Eng, P., 2009. Inequality in Indonesia: what can we learn from top incomes? J. Publ. Econ. 93 (1–2), 209–212.
Lindert, P.H., Williamson, J.G., 1982. Revising England's social tables 1688–1812. Explor. Econ. Hist. 19 (4), 385–408, Elsevier.
Marlier, E., Atkinson, A.B., Cantillon, B., Nolan, B., 2007. The EU and social inclusion: facing the challenges. J. Common Market Stud. 45 (2), 518–533.
Marx, I., Nelson, K., 2013. Minimum Income Protection in Flux. Palgrave Macmillan, Basingstoke.
Meyer, B.D., Sullivan, J.X., 2012a. Identifying the disadvantaged: official poverty, consumption poverty, and the new Supplemental Poverty Measure. J. Econ. Perspect. 26 (3), 111–136.

Meyer, B.D., Sullivan, J.X., 2012b. Winning the war: poverty from the great society to the great recession. Brookings Papers on Economic Activity 45 (2), 133–200, Economic Studies Program, The Brookings Institution.
Morelli, S., 2014. Banking crises in the US.: the response of top income shares in historical perspective. CSEF Working Papers N 359, Centre for Studies in Economics and Finance (CSEF), University of Naples, Italy.
Moriguchi, C., Saez, E., 2008. The evolution of income concentration in Japan, 1886–2005: evidence from income tax statistics. Rev. Econ. Stat. 90 (4), 713–734.
Newman, K.S., O'Brien, R., 2011. Taxing the Poor: Doing Damage to the Truly Disadvantaged. University of California Press, Berkeley.
OECD, 2008. Growing Unequal? Income Distribution and Poverty in OECD Countries. OECD Publishing, Paris.
OECD, 2011. Divided We Stand: Why Inequality Keeps Rising. OECD Publishing, Paris.
OECD, 2012. Quality review of the OECD database on household incomes and poverty and the OECD earnings database, Part I. 20 December 2012 at, http://www.oecd.org/els/soc/income-distribution-database.htm, and http://www.oecd.org/els/soc/OECDIncomeDistributionQualityReview_PartI.pdf; Table 3.
OECD, 2013. Crisis Squeezes Income Puts Pressure on Inequality and Poverty. OECD Publishing, Paris.
Orshansky, M., 1965. Counting the poor: another look at the poverty profile. Soc. Secur. Bull. 28 (1), 3–29.
Palma, J.G., 2011. Homogeneous middles vs. heterogeneous tails, and the end of the 'inverted-u': it's all about the share of the rich. Dev. Change 42 (1), 87–153.
Pareto, V., 1897. Cours d'économie politique, Lausanne and Paris, Rouge and Pichon. Reprinted in: Oeuvres Complètes. Bousquet, G.-H., Busino, G. (Eds). Genève: Librairie Droz (1964).
Piachaud, D., 1987. Problems in the definition and measurement of poverty. J. Soc. Policy 16 (2), 147–164.
Piketty, T., 2001. Les Hauts Revenus en France au Xxe Siècle: Inegalites et Redistributions 1901–1998. Grasset, Paris.
Piketty, T., 2003. Income inequality in France, 1901–1998. J. Polit. Econ. 111 (5), 1004–1042.
Piketty, T., Saez, E., 2003. Income inequality in the United States, 1913–1998. Q. J. Econ. 118 (1), 1–39.
Piketty, T., Saez, E., 2006. The evolution of top incomes: a historical and international perspective. Am. Econ. Rev. 96 (2), 200–205.
Piketty, T., Saez, E., 2012. Top incomes and the Great Recession: recent evolutions and policy implications. In: Draft for the Thirteenth Jacques Polak Annual Research Conference, Washington, DC.
Piketty, T., Saez, E., Stantcheva, S., 2012. Optimal taxation of top labor incomes: a tale of three elasticities, NBER Working Paper No. 17616, November 2011, revised October 2012.
Rainwater, L., 1990. Poverty and equivalence as social constructions. Luxembourg Income Study Working Paper #91, Center for Policy Research, Syracuse University, Syracuse, NY.
Ravallion, M., 2012. Poor, or just feeling poor? On using subjective data in measuring poverty. Policy Research Working Paper Series 5968, The World Bank, Washington, DC.
Ravallion, M., 2014. The idea of antipoverty policy. In: Atkinson, A.B., Bourguignon, F. (Eds.), Handbook of Income Distribution. In: vol. 2B. Elsevier, (Chapter 22), December.
Ravallion, M., Chen, S., 2011a. Weakly relative poverty. Rev. Econ. Stat. 93 (4), 1251–1261.
Ravallion, M., Chen, S., 2011b. Developing world is poorer than we thought, but no less successful in the fight against poverty. Q. J. Econ. 125 (4), 1577–1625.
Redmond, G., 1998. Households, families, and the distribution of income. Soc. Policy Res. Centre Newslett. 71 (1), 4–5.
Reynolds, A., 2007. Has U.S. income inequality really increased? Policy Anal. 586, 1–24.
Ringen, S., 1985. Toward a third stage in the measurement of poverty. Acta Sociol. 28, 99–113.
Roine, J., Waldenström, D., 2008. The evolution of top incomes in an egalitarian society: Sweden, 1903–2004. J. Publ. Econ. 92 (1–2), 366–387.
Roine, J., Waldenströmm, D., 2012. On the role of capital gains in Swedish income inequality. Rev. Income Wealth 58 (3), 569–587.
Rowntree, B.S., 1901. Poverty: A Study of Town Life. Macmillan, London.

Ruggles, P., 1990. Drawing the Line: Alternative Poverty Measures and Their Implications for Public Policy. Urban Institute Press, Washington, DC.
Ryscavage, P., 1995. A surge in growing income inequality? Mon. Labor Rev. 118 (8), 51–61.
Saez, E., 2013. Striking it richer: the evolution of top incomes in the United States. at, http://elsa.berkeley.edu/∼saez/.
Saez, E., Veall, M., 2005. The evolution of high incomes in Northern America: lessons from Canadian evidence. Am. Econ. Rev. 95 (3), 831–849.
Saez, E., Slemrod, J.B., Giertz, S.H., 2012. The elasticity of taxable income with respect to marginal tax rates: a critical review. J. Econ. Lit. 50, 3–50.
Sen, A.K., 1976. Poverty: an ordinal approach to measurement. Econometrica 46, 437–446.
Sen, A.K., 1983. Poor, relatively speaking. Oxf. Econ. Papers 35, 153–169.
Sen, A.K., 1992. Inequality Reexamined. Clarendon Press, Oxford.
Short, K., 2001. Experimental Poverty Measures: 1999, P60-216. US Census Bureau, Washington, DC.
Short, K., 2012. The Research Supplemental Poverty Measure: 2011, Current Population Reports: P60-244. U.S. Census Bureau, Washington, DC. November at, http://www.census.gov/prod/2012pubs/p60-244.pdf.
Simons, H.C., 1938. Personal Income Taxation: The Definition of Income as a Problem of Fiscal Policy. University of Chicago Press, Chicago.
Slemrod, J., 1996. High-income families and the tax changes of the 1980s: the anatomy of behavioral response. In: Feldstein, M., Poterba, J.M. (Eds.), Empirical Foundations of Household Taxation. University of Chicago Press, Chicago and London, pp. 169–189.
Slesnick, D.T., 1994. Consumption, needs and inequality. Int. Econ. Rev. 35, 677–703.
Smeeding, T.M., 1982. Alternative methods for valuing selected in-kind transfer benefits and measuring their effect on poverty. US Bureau of Census Technical Paper no. 50, US Government Printing Office, Washington, DC.
Smeeding, T.M., 2006. Poor people in rich nations: the United States in comparative perspective. J. Econ. Perspect. 20, 69–90.
Smeeding, T.M., Thompson, J., 2011. Recent trends in income inequality: labor, wealth, and more complete measures of income. Res. Labor Econ. 32, 1–50.
Smeeding, T.M., Weinberg, D.H., 2001. Toward a uniform definition of household income. Rev. Income Wealth 47, 1–24.
Smeeding, T.M., O'Higgins, M., Rainwater, L., 1990. Poverty, Inequality and the Distribution of Income in an International Context: Initial Research from the Luxembourg Income Study (LIS) Project. Wheatsheaf Books, London Washington, DC: Urban Institute Press.
Smeeding, T.M., Rainwater, L., Burtless, G., 2000. United States poverty in a cross-national context. Mimeo, Center for Policy Research, Syracuse University, Syracuse, NY, August.
Townsend, P., 1979. Poverty in the United Kingdom. Penguin, Harmondsworth.
Townsend, P., 1993. The International Analysis of Poverty. Harvester Wheatsheaf, London.
UNICEF Innocenti Research Centre, 2000. A League Table of Child Poverty in Rich Nations. Innocenti Report Card 1. UNICEF, Florence. June, http://www.unicef-icdc.org.
United Kingdom Department of Social Security, 1993. Households Below Average Income. Government Statistical Service, London.
United Nations Development Programme, 1999. Human Development Report. United Nations, New York July.
Van Praag, B., 1968. Individual Welfare Functions and Consumer Behavior. North-Holland, Amsterdam.
Veall, M., 2012. Top income shares in Canada: recent trends and policy implications. Can. J. Econ. 45 (4), 1247–1272.
Wang, C., Caminada, K., 2011. Disentangling income inequality and the redistributive effect of social transfers and taxes in 36 LIS countries, Luxembourg Income Study Working Paper No. 567, July.
Williamson, J.G., Lindert, P.H., 1980. American Inequality: A Macroeconomic History. Academic Press, New York.

CHAPTER 9

Recent Trends in Inequality and Poverty in Developing Countries

Facundo Alvaredo*, **Leonardo Gasparini**[†]
[*]EMod/OMI-Oxford University, Paris School of Economics, and CONICET, Oxford, UK
[†]CEDLAS-FCE-Universidad Nacional de La Plata, and CONICET, La Plata, Argentina

Contents

Handbook of Income Distribution, Volume 2A
ISSN 1574-0056, http://dx.doi.org/10.1016/B978-0-444-59428-0.00010-2

Abstract

This chapter reviews the empirical evidence on the levels and trends in income/consumption inequality and poverty in developing countries. It includes a discussion of data sources and measurement issues, evidence on the levels of inequality and poverty across countries and regions, an assessment of trends in these variables since the early 1980s, and a general discussion of their determinants. There has been tremendous progress in the measurement of inequality and poverty in the developing world, although serious problems of consistency and comparability still remain. The available evidence suggests that on average the levels of national income inequality in the developing world increased in the 1980s and 1990s and declined in the 2000s. There has been a remarkable fall in income poverty since the early 1980s, driven by the exceptional performance of China over the whole period and the generalized improvement in living standards in all the regions of the developing world in the 2000s.

Keywords

Inequality, Poverty, Income, Consumption, Developing countries

JEL Classification Codes

D31, I32

9.1. INTRODUCTION

Poverty and inequality are certainly among the main concerns in the developing world. A typical developing country is characterized by high levels of material deprivation and large dispersion in individual well-being, at least when compared to a typical high-income economy. Fighting poverty and minimizing the unjust inequalities are top priorities in the developing world. The United Nations, in the famous declaration of the Millennium Development Goals (MDGs), proposed as target number 1 to halve income poverty from 1990 to 2015. The reduction of inequality does not occupy the same privileged position in the agenda, but few would not list it as a central social concern.

Whereas Chapter 8 of this Handbook deals with poverty and inequality in advanced economies, this chapter documents patterns and changes in the developing countries.

There is no need to argue about the relevance of including a separate chapter in the Handbook: The developing world is home of 85% of total world population and bears levels of poverty and inequality far higher than in the rich nations. Whereas in a typical developing economy the share of people striving to survive with less than $2 a day is more than 30%, that share is close to zero in the industrialized countries. In fact, on this basis poverty is an issue exclusively of the developing world. The differences in income inequality are presumably also large, although the comparisons are hindered by the fact that national household surveys typically capture income in developed countries and consumption expenditures in developing ones.

High poverty and inequality are pervasive characteristics of the developing world; however, they are not immutable features of these economies. There is convincing evidence pointing to a robust decline in the levels of absolute income poverty over the last decades and substantial progress in the reduction of deprivation in various nonmonetary dimensions—education, health, sanitation, and access to infrastructure. Changes in income inequality have been much less clear, as relative inequality has risen in some countries and fallen in others. In fact, the evidence suggests that on average the developing countries are today (2014) somewhat more unequal than three decades ago.

This chapter reviews the empirical evidence on the levels and trends in income inequality and poverty in developing countries. We focus the analysis on the income/consumption approximations to welfare; in particular the chapter deals mainly with relative inequality across individuals in household consumption expenditures per capita, and with absolute poverty defined over that welfare variable, and considering alternative international lines defined in U.S. dollars adjusted for purchasing power parity (PPP). This choice is restricted by space limitations and does not imply ignoring that a general assessment of poverty and inequality should also include other nonmonetary dimensions (e.g., health, education) and other monetary variables (e.g., wealth). Other chapters in the Handbook contribute to fill those gaps.

The analysis in this chapter is mostly focused on inequality and poverty within countries and not within supranational regions or in the world.[1] Although issues of global inequality are increasingly relevant, inequality is still primarily a national concern. People are generally worried about inequality mainly in their countries, and public policies are typically aimed at reducing disparities among individuals within national boundaries.

The empirical evidence shown in this chapter is drawn from the academic literature, regional and country papers, and open-access databases, in particular the PovcalNet project developed in the World Bank. Although most of the evidence is based on statistics obtained from national household surveys, we also report results from tax records (the World Top Incomes Database, WTID) and international surveys (the Gallup World Poll) to illustrate some issues. Even though the main purpose of the chapter is to present basic

[1] Global inequality is analyzed in Chapter 11 of this volume.

evidence on levels and trends, we also briefly review the main discussions on determinants of recent changes in inequality and poverty.

The rest of the chapter is organized as follows. In Section 9.2 we briefly characterize the economies in the developing world and discuss the data sources and some measurement issues. The following two sections are assigned to the main topic in this volume—inequality. In Section 9.3 we document the levels of income inequality in the developing world, and in Section 9.4 we summarize the evidence on trends since the early 1980s. The next two sections repeat the sequence for poverty: Section 9.5 compares levels across countries, and Section 9.6 summarizes trends and discusses the evidence at the regional level.[2] Section 9.7 closes with a summary and some final remarks.

9.2. THE DEVELOPING WORLD: CHARACTERIZATION AND DATA

In this section we briefly characterize the economies of the developing world and review the sources of data to measure and analyze income poverty and inequality.

9.2.1 Developing Countries

The division between developed and developing countries is a helpful simplification that can be done in different arbitrary ways. In this chapter we follow the World Bank's main criterion based on gross national income (GNI) per capita: Developing countries are those with per capita GNI below a certain nominal threshold (US$ 12,276 in 2011). These nations are usually classified into six geographical regions: East Asia and Pacific (EAP), Eastern Europe and Central Asia (ECA), Latin America and the Caribbean (LAC), Middle East and North Africa (MENA), South Asia (SA), and sub-Saharan Africa (SSA). The Appendix includes a list of all the developing countries in each region with their populations.[3] The developing countries cover almost 75% of the total land area in the world and represent 85% of the total population. Table 9.1 summarizes some basic demographic and economic statistics.

According to these indicators Eastern Europe and Central Asia is the most developed region in the group: per capita GNI is almost twice the mean for the developing world, and the Human Development Index (HDI) is significantly higher. Latin American and the Caribbean ranks second, and Middle East and North Africa third. Although economic growth in Asia has been remarkable in the last decades, per capita GNI and other development indicators are on average still below the mean of the developing world.

[2] The separate treatment of inequality and poverty is somewhat artificial, as they are just two characteristics of the same income distribution. However, and despite some possible overlapping and duplications, we prefer to follow most of the literature and discuss both concepts separately.

[3] In this chapter we include emerging economies as part of the developing world, a decision that implies some overlapping with Chapter 8. In the period under analysis, some countries graduated from the set of developing countries; to avoid selection bias we do not drop them from the analysis.

Table 9.1 Population, GNI per capita and Human Development Index, 2010
Developing countries, by region

			GNI per capita		
	Countries	Population (millions)	PPP	Atlas method	HDI
Developing countries	153	5,840	7,023	4,291	0.608
East Asia and Pacific	24	1,961	4,911	2,992	0.619
Eastern Europe and Central Asia	30	478	12,558	7,815	0.751
Latin America and the Caribbean	31	584	9,789	6,433	0.706
Middle East and North Africa	13	331	6,462	3,647	0.636
South Asia	8	1,633	3,429	1,704	0.535
Sub-Saharan Africa	47	853	3,288	1,798	0.450
Developed countries	62	1,055	37,303	38,818	0.857
Total	216	6,894	15,682	14,181	0.663

Source: Population is taken from the United Nations Demographic Yearbook. Gross National Income (GNI) per capita in international dollars adjusted for purchasing power parity (PPP), and in current US$ (Atlas method) are taken from World Development Indicators. The Human Development Index (HDI) is from the UNDP Human Development Report. GNI and HDI are unweighted averages across countries.

South Asia is significantly less developed than East Asia and the Pacific. Sub-Saharan Africa is the poorest and least developed region of the world. The mean of the national per capita GNIs in that region is less than 50% of the developing world mean and less than 10% of the mean of the industrialized economies.

9.2.2 Data Sources

National household surveys are the main source of information for distributive analysis. Because one of the central goals of these surveys is measuring living standards, they typically include questions to construct a monetary proxy for well-being: income and/or expenditures on consumption goods. Although some developing countries started to implement national household surveys after World War II, it is only recently that governments engaged in programs of regularly collecting information through household surveys, often with the help of some international organization. Distributive statistics for the developing world are rare before the 1970s and reasonably robust only from the 1990s on. There has been a remarkable increase in the availability of national household surveys over the last decades. A chapter like this one, that includes a broad assessment of income inequality and poverty in developing countries, could hardly have been written two decades ago and is a sign of the huge progress made on data collection. However, as we discuss later, data limitations are still stringent and allow only a still blurred picture of inequality and poverty.

The databases for international distributive analysis can be classified into two groups: those that produce statistics with microdata from surveys or administrative records, and those that collect, organize, and report summary measures. The former group includes the World Bank's PovcalNet, the Luxembourg Income Study, the World Income Distribution database, the World Top Incomes Database, and some regional initiatives. The second one includes the seminal work by Deininger and Squire (1996) and its follow-up—the WIDER's World Income Inequality Database, the *All the Ginis* database, and some other projects.

The main source of information for poverty and inequality analysis at a large international scale in the developing world is the World Bank's PovcalNet, a compilation of distributive data built up from national household surveys, generally fielded by national statistical offices. PovcalNet, used for the World Bank's World Development Indicators, includes statistics constructed mostly from household survey microdata and in some few countries from grouped tabulations. At the moment of writing this database includes more than 850 surveys from almost 130 countries, representing more than 90% of the population of the developing world, spanning the period 1979–2011. The website of PovcalNet provides public access to data to generate estimates for selected countries and alternative poverty lines from grouped data.[4] Martin Ravallion and Shaohua Chen, the developers of PovcalNet, have produced several papers exploiting the data set (Chen and Ravallion, 2001, 2010, 2012; Ravallion and Chen, 1997). This project has been increasingly influential in shaping the assessment of inequality, and in particular poverty, in the developing world by researchers and policy practitioners. It is, for instance, the source used to monitor the poverty-reduction goal of the MDGs. This chapter draws heavily on statistics computed in the PovcalNet project.

Some regional initiatives aimed at estimating social statistics from harmonized household survey microdata are useful to study distributive issues in specific geographic areas and as sources of information for world databases. For instance, the *Socioeconomic Database for Latin America and the Caribbean* (SEDLAC), jointly developed by CEDLAS at Universidad Nacional de La Plata (Argentina) and the World Bank's LAC poverty unit, includes distributive and labor statistics for LAC constructed using consistent criteria across countries and years. *BADEINSO*, developed by the United Nations´ ECLAC, is also a large and good-quality database on social variables in LAC. In Eastern and Central Europe the *World Bank ECA database* includes statistics for 28 countries since 1990 computed from direct access to household surveys. The *Household Expenditure and Income Data for Transitional Economies* developed by Branko Milanovic at the World Bank is the predecessor of that database. Milanovic has also built the *World Income Distribution* (WYD) database, which includes data for five benchmark years (1988, 1993, 1998, 2002, and 2005) for 146 countries, 75% obtained from direct access to household surveys. The data set has

[4] Statistics are derived from the estimation of a general quadratic and a beta Lorenz curves from grouped data. Shorrocks and Wan (2008) propose an algorithm that reproduces individual data from grouped statistics with a higher degree of accuracy.

been used in several studies to compute global inequality (Milanovic, 2002, 2005, 2012). The *Luxembourg Income Study* (LIS), described in Chapter 8 of this volume, includes distributive information computed from household survey microdata for developed countries. LIS also reports statistics for several transitional economies in Eastern Europe and recently has added some developing countries in Latin America (Brazil, Colombia, Guatemala, Mexico, Peru, and Uruguay).

The growth in the availability of distributive statistics stimulated efforts to gather and organize them. Deininger and Squire (1996) put together a large data set of quintile shares and Gini coefficients for most countries since World War II taken from different studies and national reports.[5] This panel database, which greatly promoted the empirical study of the links between inequality and other economic variables, was updated and extended by the UNU/WIDER-UNDP *World Income Inequality Database* (WIID; WIDER, 2008).[6] The WIID database includes Gini coefficients, quintile and decile shares, and the income shares of the top 5% and bottom 5%. The information is drawn from very different sources, which raises comparability concerns.[7] To provide guidance in the use of the database, ratings are given to the observations, based on the survey quality, the coverage, and the quality of the information provided by the original source. The *SWIID* is an effort to identify reasonably comparable information in WIID (Solt, 2009).[8]

The *All the Ginis* database, assembled also by Branko Milanovic, is a compilation and adaptation of Gini coefficients retrieved from five data sets: LIS, SEDLAC, WYD, the World Bank ECA database, and WIID. Besides gathering all the information in a single file, the *All the Ginis* database is useful as it provides information on the welfare concept and recipient unit to which the reported Gini refers, facilitating the comparisons.

The *Chartbook of Economic Inequality*, assembled by Atkinson and Morelli (2012), presents a summary of evidence about changes in economic inequality (income/consumption, earnings, and wealth) in the period from 1911 to 2010 for 25 countries. The information drawn from household surveys for the seven countries in the developing world included in the database (Argentina, Brazil, India, Indonesia, Malaysia, Mauritius, and South Africa) starts in the 1950s.

[5] The Deininger and Squire data set was preceded by several earlier collections by the United Nations agencies, the World Bank, ILO, and others. See, for example, Paukert (1973), Jain (1975), and the references in Atkinson and Brandolini (2001).

[6] WIID was initially compiled over 1997–1999 for the UNU/WIDER-UNDP project "Rising Income Inequality and Poverty Reduction: Are They Compatible?" directed by Giovanni Andrea Cornia.

[7] Analyzing the Deininger and Squire data set, Atkinson and Brandolini (2001) conclude that "users could be seriously misled if they simply download the *accept* series (i.e., the 'high-quality' subset)". Although WIID implies a significant improvement from the original DS data set, a similar word of caution applies.

[8] SWIID should also be reviewed critically. In many cases it requires a case-by-case analysis, which is simply a sign that much effort is still needed in putting together comparable statistics. As it is based on secondary data sets, external problems are inadvertently incorporated.

All the data sets mentioned earlier are based on data from national household surveys.[9] Even when they are the best available source of information for distributive analysis, household surveys are plagued with problems for international comparative studies because, among other reasons, the questionnaires and the procedures to compute income/consumption variables differ among countries and frequently also within a country over time.[10] Some surveys inquire about income and others about consumption, some capture net income and some gross income, in some cases variables are reported on a weekly basis and in others on a monthly basis, items as the imputed rent for owner-occupied housing are included in some surveys and ignored in others.[11] Even in those projects that make explicit efforts to reduce these differences, comparability issues persist, as problems rooted in differences in questionnaires are difficult to be completely overcome. These limitations are well recognized in the literature. Chen and Ravallion (2012) state that "... there are problems that we cannot deal with. For example, it is known that differences in survey methods (such as questionnaire design) can create non-negligible differences in the estimates obtained for consumption or income." In a survey of global income inequality, Anand and Segal (2008) share those concerns.

There are some alternatives to reduce the comparability problems, although they all come at a price. Gallup conducts a survey in nearly all nations in the world with almost exactly the same questionnaire. The *Gallup World Poll* is particularly rich in self-reported measures of quality of life, opinions, and perceptions, but it also includes basic questions on demographics, education, and employment, and a question on household income. In principle, the Gallup World Poll allows a distributive analysis in nearly all the countries in the world based on the same income question. The downside is that measurement errors may be very large when reported income is based only on one question and with sample sizes of just around 1000 observations per country.[12]

The *Estimated Household Income Inequality* (EHII) data set produced by the University of Texas Inequality Project is based on UTIP-UNIDO, a global data set that calculates industrial pay-inequality measures for 156 countries from 1963 to 2003, using the between-groups component of a Theil index, measured across industrial categories in the manufacturing sector (Galbraith and Kum, 2005). Specifically, EHII consists on

[9] The exception is the Chartbook of Economic Inequality, which uses a range of sources, including tax data, that in some cases allows the analysis to go back much further than with household survey data.

[10] Some of these issues are also addressed in Chapter 11 of this Handbook.

[11] In addition, the typical problems of underreporting and selective compliance are negligible in some cases and endemic in others. See Deaton (2003, 2005) and Korinek et al. (2006).

[12] Gasparini and Gluzmann (2012) compare basic statistics drawn from the Gallup Poll with those computed from the national household surveys of the LAC countries for year 2006 and conclude that in most countries statistics from the Gallup Poll, including income poverty and inequality, are roughly consistent with those from national household surveys.

estimates of gross household income inequality computed from an OLS regression between the Deininger and Squire (DS) inequality measures and the UTIP-UNIDO manufacturing pay-inequality measures.[13] Although in principle the use of industrial pay information could lend some homogeneity into the comparisons, it should be stressed that because the underlying data do not refer to individuals and then have no distributive content, the methodology could be seen just as an extension of DS.

9.3. INEQUALITY: LEVELS

In this section we present results regarding the level of inequality in the developing countries, deferring to the next section the discussion of the trends. In most of the section we measure inequality computed over the distribution of household consumption per capita, using data from PovcalNet.[14] Consumption is usually regarded as a better measure of current welfare than income on both theoretical and practical grounds, especially in developing countries (Deaton and Zaidi, 2002). As it is usual in this literature, we frequently refer to *income inequality*, despite the fact that statistics are constructed over the distribution of consumption expenditures.

As discussed earlier, this chapter is mainly focused on within-country inequality, so welfare disparities are measured among individuals living within national boundaries. Although globalization is increasingly raising global inequality concerns, inequality remains mainly a national matter. This view also leads us to mostly document *unweighted* statistics of inequality measures across countries, a practice that is consistent with the typical cross-country approach in the development literature. Weighting by population would imply an assessment of inequality in a region or in the world strongly affected by some highly populated countries, such as China, India, and Indonesia in Asia, or Brazil and Mexico in Latin America, and almost ignoring the situation in other less-populated nations. Having said that, because the decision of taking each political entity as a unit in the analysis is certainly debatable, we show some results using both unweighted and population-weighted statistics.[15]

[13] The regression typically includes controls for the source of information in the inequality data (income/expenditure, gross/net, and household/per capita measures) and for the share of manufacturing employment in total employment.

[14] The drawbacks of computing inequality in the distribution of consumption or income per capita to measure distributive justice have been widely acknowledged. Among other limitations, it is a one-dimensional approach, it is focused on results not opportunities, it ignores the value of publicly provided goods such as education and health services, and it adopts a simple adjustment for demographics ignoring intrahousehold inequality, economies of scale, and differences in needs (Ferreira and Ravallion, 2009). However, extending inequality measurement to alleviate these limitations in a way that keeps international comparisons feasible has been proved difficult.

[15] See some arguments on this debate in Bourguignon et al. (2004).

9.3.1 Inequality in the Developing Countries

We start by comparing inequality levels across developing countries based on the Gini coefficient for the distribution of household consumption per capita for year 2010, computed in PovcalNet mostly from household survey microdata. Other inequality measures are highly correlated with the Gini coefficient. For instance, in PovcalNet and WIID data sets the Pearson and Spearman correlations of the Gini and several extreme inequality measures (e.g., the 90/10 and 80/20 income-share ratios) exceed 0.9.

PovcalNet includes information for the distribution of per capita consumption expenditures, except in almost all Latin American and a few Caribbean countries, for which income inequality statistics are reported. In the analysis that follows, we adjust the income Gini coefficients in that region to reflect the gap between income and consumption inequality estimates. Specifically, we selected seven Latin American countries with household surveys that include reasonably good consumption and income data in several years[16]: On average the ratio of the consumption/income Ginis is 0.861 (standard deviation of 0.046). We apply that coefficient to the 22 Latin American and Caribbean countries with income data to approximate their consumption Ginis.[17,18]

In most cases the observations correspond to year 2010, or adjacent years. However, some countries are lacking a recent household survey (or it was dropped due to quality concerns). In fact, in 24 countries the survey used to estimate inequality in 2010 was carried out between 2000 and 2005, whereas in 6 cases (5 of them in the Caribbean) the observation corresponds to the 1990s. With that caveat in mind, the PovcalNet data set has relatively recent distributive information for 82% of the countries in the developing world, representing 97% of its total population (see Table A.1 in the appendix). The country coverage across regions is heterogeneous. In East Asia and Pacific PovcalNet includes 12 out of the 24 developing countries, which nonetheless represent 96% of the total population of the area. The coverage in Eastern Europe and Central Asia is almost complete, lacking information only for Kosovo. In LAC the coverage is complete in continental Latin America, but weak in the Caribbean. Anyway, countries with information represent 98% of the total population in LAC (the main missing country in terms of population is Cuba). The data set in Middle East and North Africa does not include information for Lebanon and Libya, which represent only 3% of the MENA population. In South Asia the only country missing is Afghanistan, whereas in sub-Saharan Africa

[16] The countries are Argentina, Costa Rica, Ecuador, Mexico, Nicaragua, Panama, and Peru.

[17] We decided to apply the same coefficient to all LAC countries after failing to find significant regularities between the ratio consumption/income Ginis and other observable variables for the seven countries in the sample. World Bank (2006a) reports consumption and income Ginis in four Latin American countries; the mean ratio of the Ginis is 0.81. The value is somewhat lower (0.77) for the eight non-LA countries in the sample.

[18] We also tried an additive adjustment, instead of a multiplicative one, with no significant changes in the results.

there is information for 42 out of the 47 countries, representing 95% of the population, although in some cases the information is rather old.

Figure 9.1 displays the range of Gini coefficients for 122 countries around year 2010, ranking from the least unequal (Ukraine, 25.6) to the most unequal economy (South Africa, 63.1).[19] The mean value is 39.8, and the median is 39.2. More than half of the observations are in the range [35, 45]. Only seven Eastern Europe countries have Ginis below 30, and five sub-Saharan African countries have Ginis higher than 55. The population-weighted mean is less than one point lower than the simple mean (39.1), a result affected by the relatively low level of inequality in populous India and Indonesia (China has a Gini somewhat higher than the world mean). Figure 9.1 shows the position of some of the most populated countries: Brazil has high inequality levels, China and Russia intermediate values, and India and Indonesia relatively low levels in the context of the developing world.

The variability of Gini coefficients across countries is large compared to the changes within countries over time, at least for the period for which we have more robust information (since the early 1980s). Li et al. (1998) find in the Deininger and Squire data set that 90% of the total variance in the Gini coefficient is explained by variation across

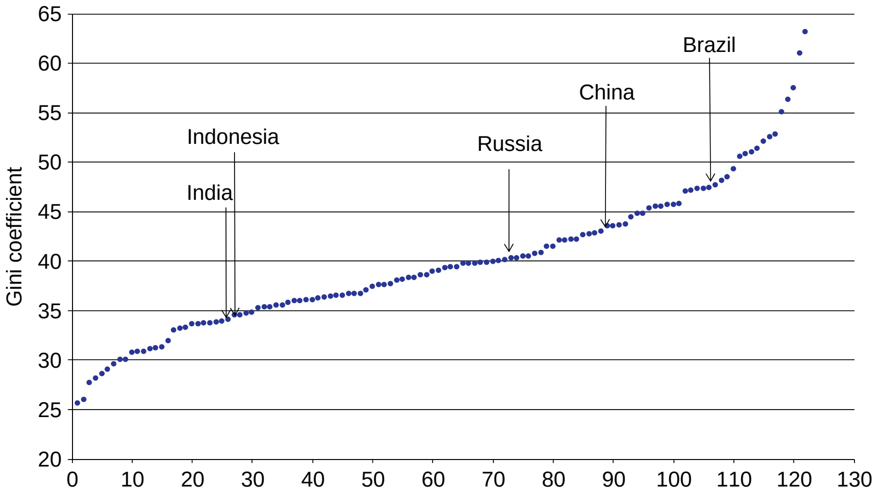


Figure 9.1 Gini coefficients for the distribution of household consumption per capita. Developing countries, 2010. *Note: Countries sorted by their Gini coefficients. Source: Own calculations based on PovcalNet (2013).*

[19] PovcalNet reports Ginis above 63.1 for Comoros and Seychelles, two small island countries in the Indian Ocean. However, the results are not well established. For instance, the reported Gini in Seychelles was 42.7 in 2000 and 65.8 in 2007, a highly implausible change in just 7 years.

countries, whereas only a small percentage is accounted for by variation over time. From this observation Li et al. (1998) conclude that inequality should be mainly determined by factors that differ substantially across countries, but tend to be relatively stable within countries over time. We find a similar result in a panel of developing countries from 1981 to 2010 (PovcalNet data): 88.5% of the variance in that panel is accounted for by variation across countries.

The inequality rankings are relatively stable over time. The Spearman-rank correlation coefficient for the Ginis in 1981 and 2010 is 0.68, whereas it rises to 0.74 for 1990 and 2010, both significant at 1%. The last decades witnessed enormous economic, social, and political changes in the developing world, but, although the income distributions have been affected with various intensities, the world inequality ranking has not changed much, a fact that suggests the existence of some underlying factors that are stronger determinants of the level of inequality.

In Figure 9.2 developing countries are grouped in regions. Sub-Saharan Africa is the geographic area that includes countries with the highest inequality levels, but it is also the region with the highest dispersion, possibly in part due to measurement errors (Table 9.2). Although eight out of the 10 highest Gini coefficients belong to sub-Saharan African countries, and the arithmetic mean of the Gini coefficient is the highest in the world, the median is lower than in Latin America.

Latin America and the Caribbean has been typically pointed out as the most unequal region in the world. Deininger and Squire (1996), for instance, stated that their data set confirm the "familiar fact that inequality in Latin America is considerably higher than in

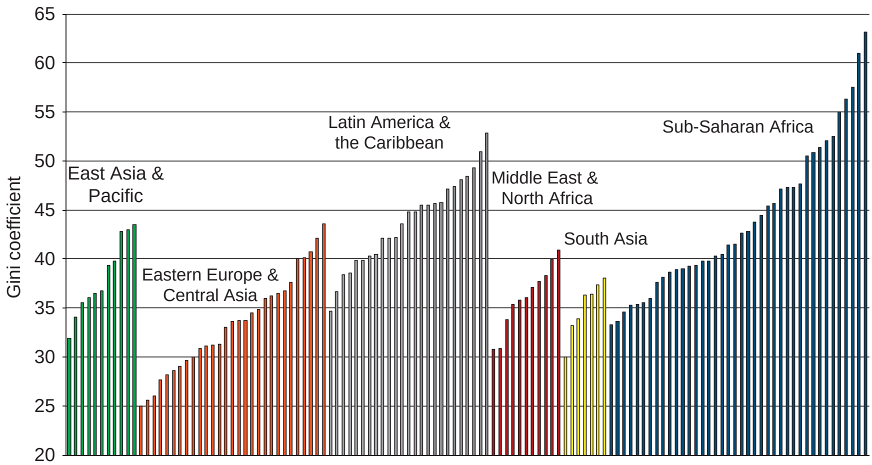


Figure 9.2 Gini coefficients for the distribution of household consumption per capita. Developing countries, 2010. *Note: Each bar represents a country in a given geographic region of the developing world. Source: Own calculations based on PovcalNet (2013).*

Table 9.2 Gini coefficients for the distribution of household consumption per capita
Developing countries, 2010

	Mean	Median	Coef. Var.	Min.	Max.
East Asia and Pacific	38.1	36.7	0.101	31.9	43.5
Eastern Europe and Central Asia	33.6	33.7	0.144	25.6	43.6
Latin America and the Caribbean	43.8	44.8	0.104	34.7	52.8
Middle East and North Africa	36.0	36.1	0.091	30.8	40.9
South Asia	35.0	36.3	0.081	30.0	38.1
Sub-Saharan Africa	44.4	42.1	0.175	33.3	63.1
Developing countries	39.8	39.2	0.181	25.6	63.1

Note: Unweighted statistics.
Source: Own calculations based on PovcalNet (2013).

the rest of the world."[20] This type of assessment, however, is usually made combining income Ginis for LAC with consumption Ginis for other regions and/or ignoring sub-Saharan Africa. With the adjustment mentioned earlier to take the consumption/income gap into consideration (factor 0.861), we find that the mean Gini for LAC is 43.8, slightly lower than in SSA (44.4), but the median is higher (44.8 in LAC and 42.1 in SSA). To reach the result of a higher mean Gini in LAC than in SSA, we would need an adjustment parameter higher than 0.92; such value is larger than what we estimated in all LA countries in the sample, except Mexico.

The rest of the regions in the developing world have Ginis mostly below 40. The arithmetic mean is 38.1 in East Asia and Pacific, 36.0 in Middle East and North Africa, and 35.0 in South Asia. Inequality is likely to be higher in MENA because several oil-producing countries are excluded for being high-income economies (and also for lack of information).[21] Eastern Europe and Central Asia is the region with the lowest inequality levels, with a mean Gini coefficient of 33.6. Interestingly, the dispersion measured by the coefficient of variation is higher than in the rest of the regions, except SSA.

Almost all very highly unequal countries (Gini coefficients above 50) are in sub-Saharan Africa (Table 9.3). This region, however, has a similar share of countries in the high and middle categories. In contrast, in LAC most countries have high levels of inequality, whereas in EAP, MENA, and SA most countries are in the middle-inequality group. Only ECA has economies with low inequality (Gini coefficients below 30).

The *All the Ginis* data set (ATG) includes Gini coefficients from LIS, SEDLAC, WYD, the World Bank ECA database, and WIID. We selected consumption Ginis from ATG for year 2005 or close and applied a similar adjustment as described earlier for those countries in LAC with only income Ginis. The basic results are similar to the ones

[20] See also López Calva and Lustig (2010) and Chen and Ravallion (2012).
[21] Bahrain, Kuwait, Oman, Qatar, Saudi Arabia, and United Arab Emirates are in that group. Malta and Israel are also ignored for being developed, and Lebanon and Libya are excluded for lack of information.

Table 9.3 Classification of countries by level of inequality and by region
Developing countries, 2010

	Inequality				
	Very high [50–70]	High [40–50]	Middle [30–40]	Low [20–30]	Total
East Asia and Pacific	0	3	8	0	11
Eastern Europe and Central Asia	0	5	16	7	28
Latin America and the Caribbean	2	17	6	0	25
Middle East and North Africa	0	1	10	0	11
South Asia	0	0	7	0	7
Sub-Saharan Africa	10	14	16	0	40
Total	12	40	63	7	122

Note: Countries are classified according to the value of the Gini coefficient for the distribution of household consumption per capita.
Source: Own calculations based on PovcalNet (2013).

obtained with PovcalNet data. The linear correlation coefficient for the Gini between both data sources is 0.763, whereas the Spearman rank correlation is 0.771, both significant at 1%. The Gini coefficients in ATG go from 23.1 (Czech Republic) to 62.9 (Comoros). The mean and median coincide in 40.1. Again, more than half of the observations are in the range [35, 45]. Only several Eastern European countries have Ginis below 30, whereas only four sub-Saharan African countries have Ginis higher than 55.

The evidence on inequality levels in the developing world drawn from WIID is similar. For instance, based on a sample of income Ginis for around 2005, Gasparini et al. (2013) find that the mean Gini for the six sub-Saharan African countries in the data set is 56.5, followed by Latin America (52.9), Asia (44.7), and Eastern Europe and Central Asia (34.7).[22] The linear correlation coefficient for year 2005 for the Gini coefficient in PovcalNet and WIID is 0.871, and the Spearman coefficient is 0.820.

The Luxembourg Income Study database (see Chapter 8 of this volume) covers 36 countries, including 6 in Latin America, which occupy the top places in all the income inequality rankings.[23] The mean Gini for the Eastern European countries in LIS is slightly higher than the mean for the high-income economies. Data from the World Development Indicators also suggest that inequality in the developing world is significantly higher than in the OECD high-income countries. The mean income Gini for the latter group is 32.2, which is lower than in any other region in the world.

[22] The OECD high-income countries rank as the least unequal in the world with a mean income Gini of 32.8.

[23] The LA Ginis go from 50.6 in Colombia to 43.9 in Uruguay; the most unequal non-LA country is Russia with a value of 40.8, whereas the rest of the countries in LIS go from 37 (USA) to 22.8 (Denmark).

The EHII database confirms the high inequality levels of sub-Saharan Africa and Latin America, but perhaps surprisingly, it records similar levels in South Asia and Middle East and North Africa (Gini of around 47).[24] According to this data set, inequality is relatively lower in East Asia and Pacific and Eastern Europe and Central Asia. The estimated level of the Gini coefficient is substantially lower in the developed economies; the mean is equal to 36.5.[25] The Pearson (Spearman) correlation coefficient between EHII and PovcalNet Ginis is 0.642 (0.603), lower than the resulting value when comparing PovcalNet with WIID or ATG, but still significant at 1%.

Most international databases do not provide confidence intervals for the point estimates of the distributive measures, making impossible the assessment of the statistical significance of the differences in inequality among countries. However, given that the indicators are calculated from large national household surveys, the confidence intervals are typically relatively narrow. SEDLAC provides the confidence intervals for all the Gini coefficients in Latin America. For instance, the 95% confidence interval for the income Gini was [43.9, 44.7] in Argentina 2010, [53.5, 54.0] in Brazil 2009, and [47.0, 47.9] in Mexico 2010. Differences in the point estimates of more than 1 Gini point are always statistically significant (Figure 9.3).

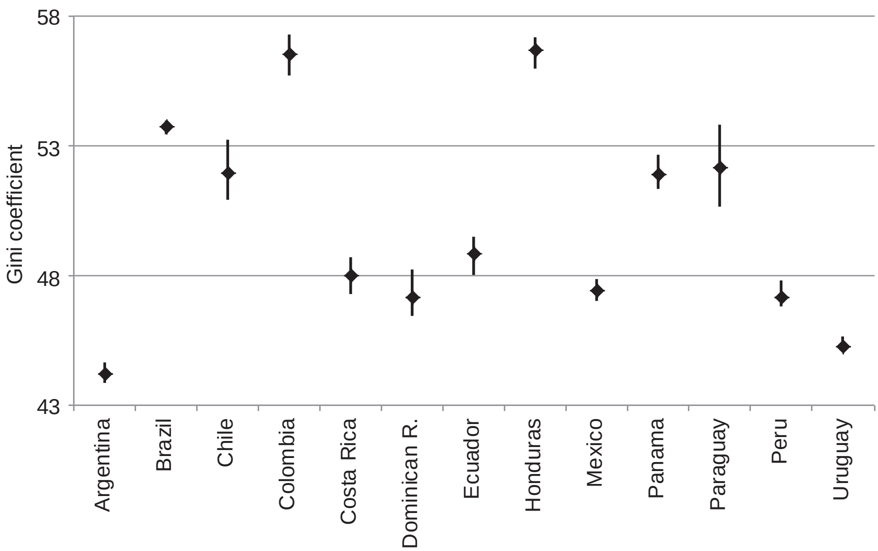


Figure 9.3 Gini coefficient and confidence intervals (95%). Distribution of household income per capita. Latin American countries, 2010. *Source: Own calculations based on SEDLAC (CEDLAS and the World Bank).*

[24] See also Galbraith and Kum (2005).

[25] This mean excludes the oil-rich Arab countries. When including these countries in the sample, the mean Gini jumps to 39.

9.3.2 Inequality Beyond the Gini Coefficient

The international databases usually allow a closer look at the distributions in the world beyond a single parameter, such as the Gini coefficient. Table 9.4 reports some basic statistics of the decile shares in 120 countries around 2010.[26] On average (unweighted) the poorest 10% of the population in a country accrues 2.6% of total consumption reported in the survey; that share climbs to 31.5% for the top 10%. In a typical developing country the aggregate consumption of the poorest 60% of the population is similar than the consumption of the top 10%.

It is interesting to notice that the coefficient of variation of the decile consumption shares across countries is decreasing up to the top decile, when it strongly rises. Countries in the world seem substantially different in the consumption share of the poor and the rich, but not in the share of the middle strata, in particular the upper-middle strata.[27]

The aggregate consumption share of deciles 5–9 is on average around 50%, and it is very stable across countries. Palma (2011) has labeled this phenomenon the *homogeneous middle*. Variability across countries is actually smaller in the upper-middle deciles (deciles 7–9). The proportion of total consumption accruing to that group is quite similar in all geographic regions of the world; it ranges from 35.9% in SSA to 37.3% in ECA. The main difference across regions lies in the share of the bottom 60% compared to those

Table 9.4 Deciles shares, distribution of household consumption per capita
Developing countries, 2010

Deciles	Mean	Std. Dev.	Coef. Var.	Min.	Max.
1	2.6	0.81	0.31	1.0	4.4
2	3.8	0.86	0.23	1.5	5.8
3	4.8	0.90	0.19	2.0	6.8
4	5.8	0.92	0.16	2.6	7.8
5	6.8	0.92	0.13	3.5	8.8
6	8.1	0.87	0.11	4.7	9.9
7	9.6	0.80	0.08	6.6	11.0
8	11.7	0.65	0.06	9.0	12.7
9	15.3	0.84	0.05	12.7	17.6
10	31.5	6.12	0.19	19.5	51.7

Note: Unweighted statistics.
Source: Own calculations based on PovcalNet (2013).

[26] Again, figures for Latin American and a few Caribbean countries are estimated based on the comparison of income and consumption microdata of seven countries in that region.

[27] This observation could be simply linked to the fact that the cumulative distribution functions of two income distributions most often cross around the middle (e.g., in a mean preserving spread) rather than at the ends of the distributions.

in the upper 10%. For example, whereas the share of deciles 7–9 in total consumption is almost the same in ECA (37.3%) and LAC (37.1%), the share of the bottom 60% is more than 7 points higher in the former (36.4% and 29.1%, respectively).

The correlation coefficients for the decile shares in total consumption provide information about the structure of the distributions across countries (Table 9.5). In a cross-country perspective, gains are highly positively correlated in the first 8 deciles; on the other hand, for decile 10 correlations are all negative and large, except with decile 9, for which the correlation is non-significant. Gains in the participation of the richest 10% are tightly linked to losses in the share of the poorest 80% of the population. The table suggests that when we move up in the ladder of countries according to the share of the bottom deciles, we expect to see gains in the lowest strata obtained mostly against the share of the upper 20% of the population (and not, for instance, against the middle strata, and in alliance with the most affluent).

9.3.3 Inequality in the Gallup World Poll

The Gallup World Poll provides new evidence on the international comparisons of income inequality, as it includes identical income and demographic questions applied to national samples in 132 countries. Of course, the reliability of the national inequality estimates in Gallup is lower than those obtained with household surveys because only one income question is used to approximate well-being, and the sample sizes are considerably smaller. However, Gluzmann (2012) finds that the correlation coefficient between the Gini coefficients computed with Gallup microdata and those reported in the World Development Indicators (WDI) that are based on per capita income is high (0.85).[28] International surveys with similar questionnaires across countries, such as the Gallup World Poll, could hardly be a substitute for household surveys as the main source for distributive analysis at the country level, but they may have a great potential for international comparisons of social variables. Future improvements in the quality of these surveys could turn them into a very valuable source for comparative international research.

Gasparini and Gluzmann (2012) use microdata from the Gallup World Poll 2006 to compute inequality in each region of the world. According to the unweighted mean of the national income Gini coefficients, Latin America is the most unequal region in the world (excluding Africa, which is not in the sample). The mean Gini in Latin America is 49.9, slightly larger than in South Asia (48.9) and Eastern Asia and Pacific (47.1). Countries in Eastern Europe and Central Asia (41.8), North America (39.2), and especially Western Europe (34.0) are the least unequal. Alternatively, regional inequality can be measured by considering each region as a single unit and computing inequality among

[28] Interestingly, the relationship between the income Ginis in Gallup and the consumption Ginis in WDI is much weaker; the linear correlation coefficient is 0.21, non-significant at 10%.

Table 9.5 Correlation coefficients across countries of decile consumption shares
Developing countries, 2010

	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
d1	1									
d2	0.9355*	1								
d3	0.8930*	0.9883*	1							
d4	0.8421*	0.9624*	0.9910*	1						
d5	0.8042*	0.9273*	0.9647*	0.9787*	1					
d6	0.7336*	0.8739*	0.9291*	0.9623*	0.9847*	1				
d7	0.6310*	0.7734*	0.8436*	0.8950*	0.9378*	0.9736*	1			
d8	0.3127*	0.4711*	0.5624*	0.6446*	0.7253*	0.8085*	0.8982*	1		
d9	−0.5793*	−0.4905*	−0.4112*	−0.3258*	−0.2389*	−0.1232	0.0527	0.4390*	1	
d10	−0.7844*	−0.9032*	−0.9452*	−0.9689*	−0.9844*	−0.9891*	−0.9650*	−0.7962*	0.118	1

*Significant at 1%.
Source: Own calculations based on PovcalNet (2013).

all individuals in that unit, after translating their incomes to a common currency—a concept usually labeled *global inequality* (see Chapter 11 of this Handbook). The global Gini in Latin America is 52.5, a value higher than in Western Europe (40.2), North America (43.8), and Eastern Europe and Central Asia (49.8), but lower than in South Asia (53.2) and Eastern Asia and Pacific (59.4). The change in the rankings between the two concepts of inequality is driven by the differences across regions in the heterogeneity among countries in terms of mean income. Gasparini and Gluzmann (2012) report that the between component in a Theil decomposition accounts for 8% of total regional inequality in Latin America and 32.4% in East Asia and Pacific.

9.3.4 Top Incomes

Until the recent developments in the literature of top incomes from tax records (Atkinson and Piketty, 2007, 2010; see also Chapter 8 in this volume), inequality research has been mostly based on household surveys, which suffer from several limitations when focusing on the upper end of the distribution. Household surveys are all but ideal for studying top shares: The rich are usually missing from surveys, either for sampling reasons or because they refuse to cooperate with the time-consuming task of completing or answering a long form. Because extreme observations are sometimes regarded as data "contamination," the rich may be intentionally excluded or top coded so as to minimize bias problems generated by presumably less-reliable outliers or to preserve anonymity. In addition, survey data present severe underreporting at the top; the richest individuals are more reluctant to disclose their incomes or have diversified portfolios with income flows that are difficult to value.

Székely and Hilgert (1999) look at surveys from eighteen Latin American household surveys and confirmed that the ten highest incomes reported are often not much larger than the salary of an average manager in the given country at the time of the survey. In general, the profile of the average individual in the top 10% of the distribution is closer to the prototype of highly educated professionals earning labor incomes, rather than capital owners. On this specific issue, the quality of statistical information coming from surveys has not improved in the last years. Consequently, the inequality that we are able to measure with household surveys can be severely affected, regarding both levels and dynamics, in those cases or periods in which an important part of the story takes place at the top.

Tax and register data are being increasingly preferred over surveys in studying distributive issues at the top. In fact, under certain conditions registry data can provide valuable information to improve survey-based estimates. Typically, incomes reported to the surveys are checked against the registers, or incomes are directly taken from administrative sources for the individuals in the sample. Even if the combination of survey and administrative data can be seen as an improvement, there remains the issue of the sampling framework for the top of the distribution.[29] In any case, statistics offices in the developing world are not exploiting register data to complement surveys yet.

The use of tax statistics is not without drawbacks. First, because only a fraction of the population files a tax return, studies using tax data are restricted to measuring top shares, which are silent about changes in the lower and middle part of the distribution. Second, tax data are collected as part of an administrative process and do not seek to address research needs; both income and tax units are defined by the tax laws and vary considerably across time and countries. Third and most important, estimates are affected by tax avoidance and tax evasion; the rich, in particular, have a strong incentive to understate their taxable incomes. These elements, which are common to all countries, become critical in the developing world, characterized by tax systems with low enforcement and multiple legal ways to avoid the tax.[30]

A number of researchers have addressed the differences in the ability of tax and survey data to represent income inequality, trying to reconcile the evidence using the two sources (see Alvaredo, 2011; Burkhauser et al., 2012 for the United States). Unfortunately, at the moment of writing only a few developing countries have made available microdata from the income tax (namely Colombia, Ecuador, and Uruguay). Alvaredo and Londoño (2013), and Alvaredo and Cano (forthcoming) show that, in contrast to survey-based results, high-income individuals are, in essence, rentiers and capital owners. This feature differs from the pattern found in several developed countries in recent decades, where it has been shown that the large increase in the share of income going to the top groups has been mainly due to spectacular increases in executive compensation and high salaries, and to a lesser extent to a partial restoration of capital incomes. Although the working rich have joined capital owners at the top of the income hierarchy in the United States and other English-speaking countries, Colombia and Ecuador remain more traditional societies where the top-income recipients are still the owners of the capital stock.

Results, even if fragmentary, confirm that incomes reported to the tax authorities are considerably higher than those captured by the surveys at the top. For instance, the share of income accrued by the top 1% in Argentina in 2007 was 8.8% using household survey data (PovcalNet) and 13.4% using income tax data (WTID). In Uruguay 2010 the shares were 8.2% and 14.3%, and in Colombia 2010 13.9% and 20.4%, respectively. Even if those numbers are not directly comparable (surveys incomes are before tax), they show

[29] If high-income individuals are not properly identified in the sample framework, comparing the incomes reported to the surveys against those in the registers one by one is only a partial improvement. In the UK, for example, the ONS scales up the surveys' incomes so that the surveys' averages match the average income in tax data.

[30] The reasons for which the rich and wealthy may be particularly dissuaded from disclosing their fortunes and incomes to authorities in the developing world may go beyond tax concerns, lest the information revealed fall into the wrong hands. Alvaredo and Londoño (2013) report that in Colombia, until recently plagued by high insecurity, anecdotal evidence suggests that during the intense political violence of the 1990s, leaked personal tax returns were used by criminal groups to target victims and kidnap for ransom.

that synthetic measures of inequality, if presented in an isolated way, hide survey-based shares that may be unrealistically low. In this sense, it could be a good practice to systematically show the inequality indexes together with the shares of the underlying top percentiles to let users judge the quality of the estimates.[31]

A natural question, which has received much attention lately, is the extent to which tax data can complement household surveys in examining the level of inequality in developing countries. Alvaredo and Londoño (2013) compare the Colombian household survey with the tax micro-data over the years 2007–2010. The total household income from the survey is 60–65% of the NAS measure of disposable income.[32] Such gap cannot be seen as an accurate measure of the total missing income in household surveys because both sources are different, but a partial explanation may well be at the top of the distribution. As a simple exercise, these authors replace all the incomes above the percentile 99 in the survey with those from tax data (net of taxes and social security contributions to render both sources comparable), under the assumption that the top 1% is poorly captured in the survey. Two elements are worth mentioning. First, the gap between the NAS figure and the survey's incomes of the bottom 99% plus the net-of-tax incomes from tax data above the percentile 99 goes down from 35–40% to 20–25%. Second, the Gini coefficient of individual incomes goes up from 55 to 61 in 2010.[33]

These findings challenge the general skepticism regarding the use of tax data from developing countries to study inequality. Such estimates should be regarded as a lower bound, to take into account the effects of evasion and underreporting. Nevertheless, they show that incomes reported to tax authorities can be a valuable source of information, under certain conditions that require a case-by-case analysis.

9.3.5 Inequality and Development

Is the level of inequality in a country associated to its development stage? In this section we take advantage of a cross section of national Gini coefficients for year 2010 to take a look at this issue. Of course, this topic is related to the long-lasting debate initiated with the seminal contributions by Lewis (1954) and Kuznets (1955), who argued that the process of industrialization would imply an inverse U pattern for inequality. However, the empirical test for the Kuznets curve requires time-series or panel data, and not just a cross

[31] Povcalnet follows this practice by providing estimates of the Lorenz curve, with varying degrees of detail depending on the country.

[32] The National Accounts-based measure of household disposable income has been defined as: balance of households' primary incomes + social benefits other than social transfers in kind − employers' actual social contributions − imputed social contributions − attributed property income of insurance policyholders − imputed rentals for owner occupied housing − fixed capital consumption − employees' social security contributions − taxes on income and wealth paid by households.

[33] These results are still approximations, as defining individual actual incomes from the Colombian tax records is not always straightforward.

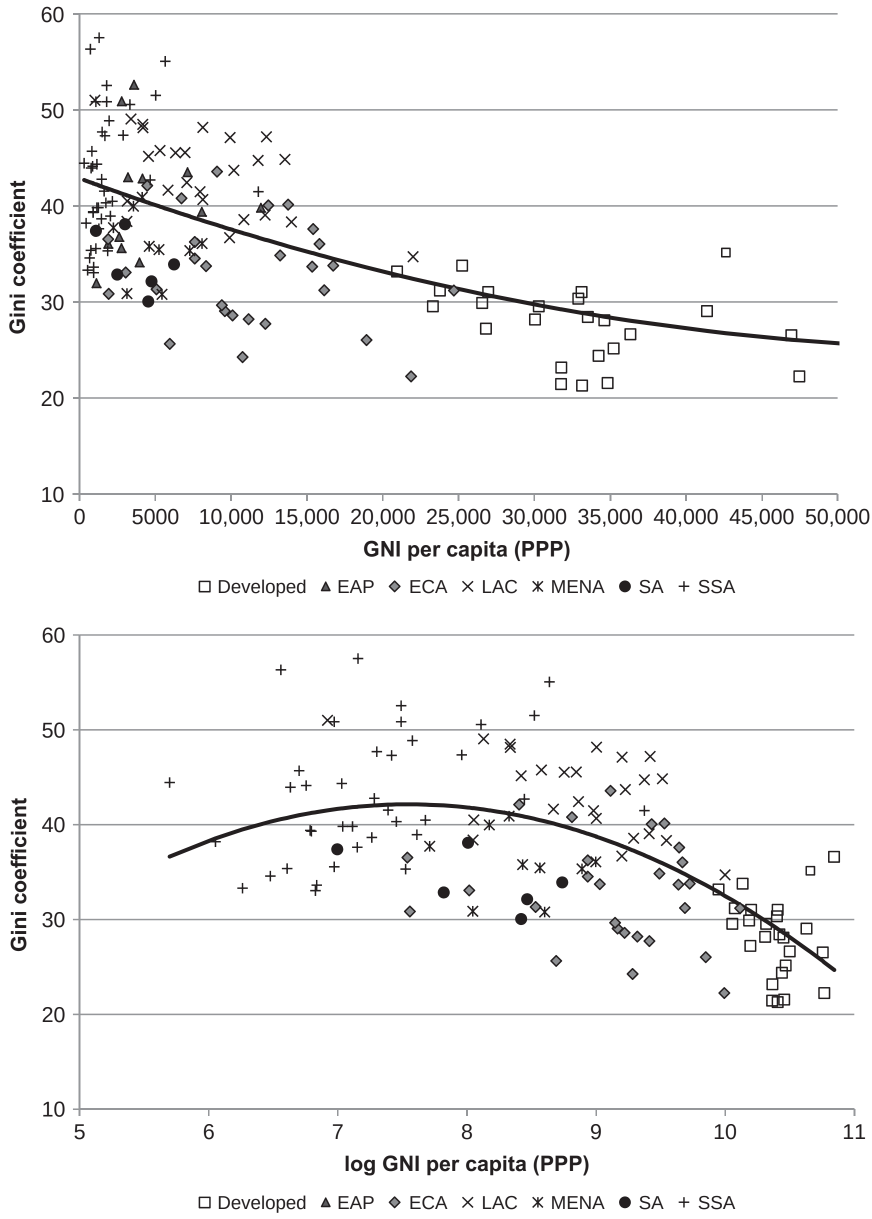


Figure 9.4 Inequality and development. Per capita gross national income (GNI) and Gini coefficient, 2010. *Source: Own calculations based on WDI and PovcalNet (2013).*

section, because it is a hypothesis about the dynamics of an economy over its development process. The causal relationship between development and inequality is the subject of a large literature that has to face numerous empirical challenges, and hence it is far from settled (see Anand and Kanbur, 1993; Banerjee and Duflo, 2003; Fields, 2002; Voitchovsky, 2009, for assessments). In this section we simply document the empirical relationship between these two variables across countries in a recent point in time without exploring the difficult issue of causality.

The first panel in Figure 9.4 plots the Gini coefficient for the distribution of consumption per capita against per capita gross national income (GNI).[34] The figure seems to

Table 9.6 Regressions of Gini coefficient on log GNI per capita and regional dummies

	All countries		Only developing countries	
	(i)	(ii)	(iii)	(iv)
log GNIpc	24.24 (9.52)★★	24.44 (4.48)★★★	18.01 (8.23)★	26.54 (6.58)★★
log GNIpc squared	−1.606 (0.552)★★	−1.409 (0.34)★★★	−1.202 (0.53)★	−1.541 (0.48)★★
Developed countries		−1.416 (2.76)		
East Asia & Pacific		7.352 (1.43)★★★		7.170 (1.62)★★★
Latin America & Caribbean		10.238 (0.53)★★★		10.157 (0.62)★★★
Middle East & North Africa		2.334 (1.28)		2.144 (1.48)
South Asia		1.705 (1.79)		1.515 (1.97)
Sub-Saharan Africa		13.749 (2.33)★★★		13.660 (2.34)★★★
Constant	−49.34 (38.69)	−72.10 (13.17)★★★	−61.67 (28.64)★★	−80.27 (20.06)★★
Observations	146	146	121	121
R-squared	0.31	0.58	0.07	0.45
Lind and Mehlum test for inverse U shape				
$\|t\|$	2.72	2.31	1.35	2.0
p-Value	0.004	0.011	0.089	0.024

Note: Robust cluster standard errors in brackets.
Omitted category: Eastern Europe and Central Asia.
Lind and Mehlum test: H_0: monotone or U shape; H_1: inverse U shape.
★significant at 10%; ★★significant at 5%; ★★★significant at 1%.

[34] The Gini for the developed countries is computed over the distribution of income per capita, and not consumption per capita, a fact that probably underestimates the slope of the curve.

reveal a decreasing relationship between inequality and development. The linear correlation coefficient between the Gini coefficient and per capita GNI is −0.56 (statistically significant at the 1% level). An inverse-U shape shows up in the second panel of Figure 9.4, when per capita GNI is presented in logs. However, the increasing segment of the curve covers only very poor sub-Saharan African countries. The relationship Gini-GNI is decreasing in the range of GNI of most countries in the world.

The results of the regressions in Table 9.6 and the Lind and Mehlum (2010) test confirm an inverse U shape for the relationship between the Gini coefficient and log GNI per capita in a cross section of countries.[35] The result seems also valid, although it becomes considerable weaker when restricting the sample to developing economies. It should be stressed that the turning points implicit in the regressions correspond to around US$ 1800, a value that is lower than the per capita GNI of most developing countries, except for some economies in sub-Saharan Africa.[36] The inclusion of regional dummies reveals that East Asian, and especially Latin American and sub-Saharan African, countries are particularly unequal, even when controlling for their levels of economic development.[37]

9.4. INEQUALITY: TRENDS

In this section we report the recent trends in income inequality in the developing countries. We start laying out the general patterns, and then dig deep into the evidence for each region. Although most of the section deals with relative inequality, we devote a section to explore patterns for absolute inequality and a section to document aggregate welfare changes.[38] We end with a brief summary of the methodologies and main issues in the debate on inequality determinants in the developing world.

9.4.1 General Changes

The available evidence suggests that on average the levels of national income inequality in the developing world increased in the 1980s and 1990s and declined in the 2000s. Using data from PovcalNet, the mean Gini for the distribution of per capita consumption

[35] It is also confirmed estimating GDP with the Atlas method and using the *All the Ginis* database.

[36] Larger measurement errors in the SSA countries may also account for the increasing segment of the curve. Also, it is possible that the econometric model is picking up the concavity of the relationship at higher income levels.

[37] The Latin American "excess inequality" is documented in Londoño and Székely (2000); Gasparini, Cruces and Tornarolli (2011a), and others.

[38] Although relative inequality measures are scale invariant, absolute measures are translation invariant. Accordingly, a general increase of x% in all incomes in the population will leave relative inequality unchanged, but imply an increase in absolute inequality.

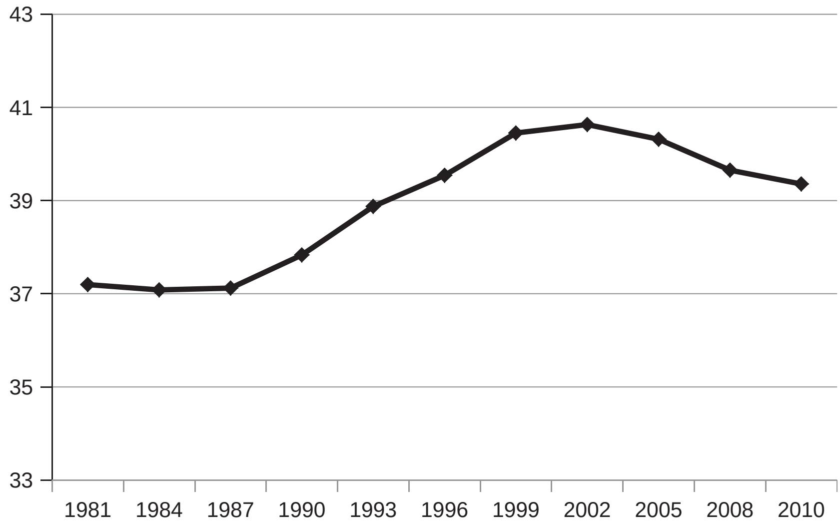


Figure 9.5 Gini coefficient. Unweighted mean for developing countries, 1981–2010. *Note: The national Gini coefficients are computed over the distribution of household consumption per capita. Source: Own calculations based on PovcalNet (2013).*

expenditures increased from 37.2 in 1981 to 39.4 in 2010 (Figure 9.5).[39] The mean was basically unchanged between 1981 and 1987,[40] then increased more than three points to reach a value of 40.5 in 1999, and from 2002 it started to fall, although slowly (from 40.6 in 2002 to 39.4 in 2010).[41]

Figure 9.6 adds to the picture the changes at different percentiles of the distribution of national Ginis. The figure makes clear that on average the changes in the last decades have not been large compared to the range over which the Gini varies across countries.[42] The picture also reveals that the growth in the mean Gini in the late 1980s and 1990s

[39] To compute changes we discard countries in PovcalNet with less than four observations over the period 1981–2010, or with observations concentrated in a narrow time-period. The sample we use for the calculations on trends include 76 countries that represent 88% of the developing world population. To build a sample in which the country composition is held constant in a few cases, Gini coefficients are imputed assuming constant inequality. Income Ginis in LAC are adjusted as explained in the previous section.

[40] This result is in part driven by the lack of information on changes in inequality over this period for several countries in the developing world. See below.

[41] The assessment of the economic salience of inequality changes over time is controversial as it involves both issues regarding the accuracy of the data and considerations on the purpose for which the inequality statistics are used. In the context of the OECD countries, Atkinson and Marlier (2010) propose applying a 2% points criterion to assess the salience of the change in the Gini. On this basis, the increase in inequality in the developing world from 1981 to 2010 is just salient.

[42] This observation does not imply that changes were of little social relevance; an increase in inequality in a given country could be small in relation to the difference with other countries, but still a major cause of concern.

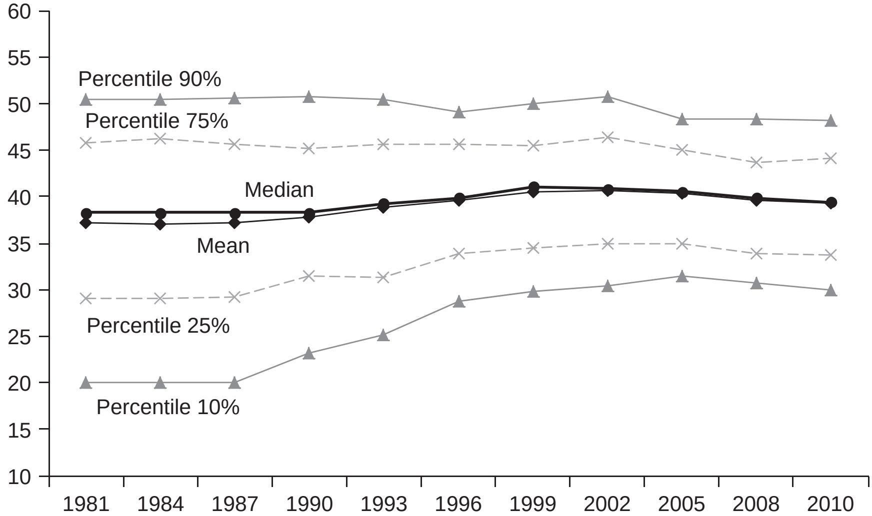


Figure 9.6 Distribution of Gini coefficients. Unweighted statistics for developing countries, 1981–2010. *Note: The national Gini coefficients are computed over the distribution of household consumption per capita. Source: Own calculations based on PovcalNet (2013).*

was mainly due to the substantial increase in the low-inequality countries, in particular Eastern Europe and Central Asia economies after the fall of communism, and also some Asian economies in the early stage of economic takeoff. Instead, the fall in the 2000s was widespread, although more intense in those countries above the median, such as those in Latin America. This observation suggests convergence in the levels of inequality in the developing economies. In fact, the standard deviation for the distribution of Gini coefficients substantially fell over time: 11.2 in 1981, 10.1 in 1990, 7.4 in 1999, and 7.2 in 2010. Countries in the developing world are still very different in terms of income inequality, but differences have become considerably smaller over the last three decades (more on convergence below).

A closer inspection of the data reveals that the result of a stable mean Gini in most of the 1980s is driven by the lack of information for several countries and by a substantial heterogeneity in the changes of those with information (Table 9.7).[43] The strong rise in the mean Gini in the 1990s is associated with a large proportion of countries with growing inequality in a framework of much improved information. The tide seems to have turned in the 2000s, when most of the countries in the sample experienced a fall in

[43] We classify countries in groups according to whether the Gini went up or down by more or less than 2.5% in a period. A change of 2.5% applied to the mean Gini in the developing world—which is around 40—represents 1 Gini point. A change of 1 point in the Gini coefficient is typically statistically significant, given the sample sizes of the national household surveys.

Table 9.7 Proportion of countries classified in groups according to the change in the Gini coefficient

	1981–1990	1990–2002	2002–2010
Fall	14.7	22.7	65.3
No change	21.3	16.0	14.7
Increase	34.7	60.0	20.0
No information	29.3	1.3	0.0
Total	100.0	100.0	100.0

Note: "Fall" includes countries where the Gini fell more than 2.5% in the period; "Increase" includes countries where the Gini rose more than 2.5%; "No change" includes countries where the Gini changed less than 2.5%; "No information" includes countries without two independent observations in each period.
Source: Own calculations based on PovcalNet (2013).

inequality. But even in this decade of widespread social improvement, the country performances in terms of inequality reduction were quite heterogeneous. In fact, in 20% of the economies of the developing world the Gini coefficient increased between 2002 and 2010, whereas in 15% of the countries the changes were smaller than 2.5%.

We find that the bulk of the countries in the sample (62%) experienced a change in the pattern of inequality around the turn of the century, from nonfalling to decreasing inequality, whereas only a few experienced a pattern of continuous increasing (15%) or decreasing (12%) disparities. In fact, an inverse-U shape for the inequality pattern is observed for many economies (45% of the sample), a fact that could be consistent with the Kuznets story of economic growth for countries located close to the curve turning point. However, we fail to find any significant correlation between the type of the inequality pattern and different measures of development and growth. The inverse-U pattern in the period 1981–2010 appears to have been common to a wide range of economies.

The growth in the population-weighted mean of the Gini coefficient across developing countries was stronger than the increase in the unweighted mean (Figure 9.7). Although the latter increased 2.2 points in the period 1981–2010, the former jumped 7.5 points. The gap between the two means shrunk from 5.4 points in the early 1980s to almost zero in the late 2000s. This pattern is mainly accounted for by the dramatic surge in income inequality in China over the period. Interestingly, the fall in the unweighted mean Gini in the 2000s does not show up in the weighted mean; although the Gini coefficient for a typical developing country significantly decreased in the 2000s, the national Gini for a typical person in the developing world did not fall.

In the rest of this section we go beyond the Gini coefficient and track changes along the distribution. In Figure 9.8 each point in a growth-incidence curve (GIC) indicates the unweighted mean across countries in the annual rate of growth of real consumption per capita (in PPP US$) for a given decile of the national distributions.[44] There is a stark

[44] The GIC depicted in Figure 9.8 is not the world growth-incidence curve, where, for instance, decile 1 would include the poorest 10% of the world population.

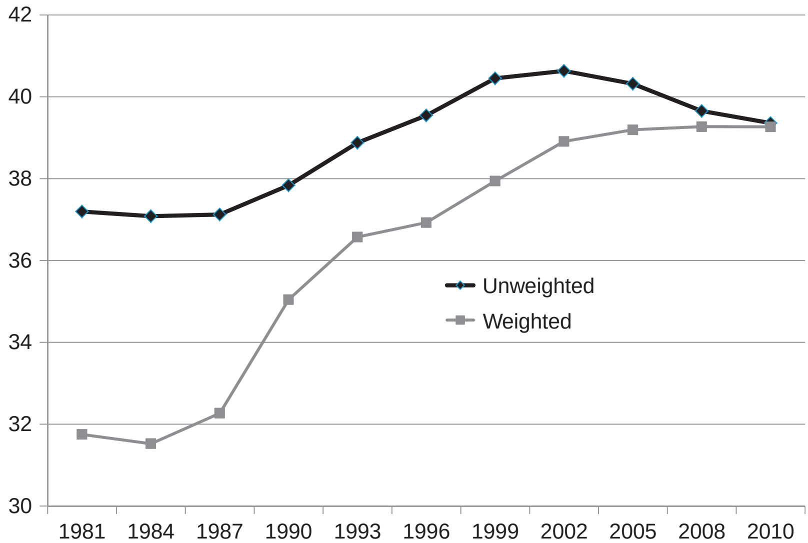


Figure 9.7 Gini coefficient. Weighted and unweighted means. Developing countries, 1981–2010. *Note: The Gini coefficients are computed over the distribution of household consumption per capita. Source: Own calculations based on PovcalNet (2013).*

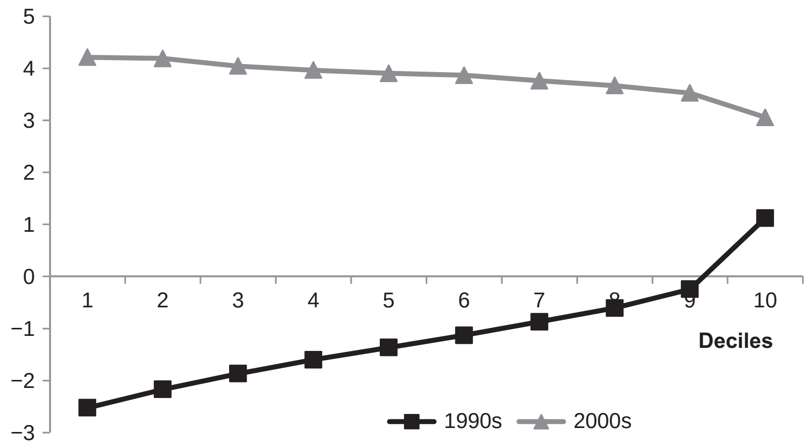


Figure 9.8 Growth-incidence curves. Annualized growth rate in consumption per capita by decile. Unweighted mean for developing countries. *Note: Annual change in consumption per capita (PPP US$). 1990s = 1990–2002, 2000s = 2002–2010. Source: Own estimates based on PovcalNet (2013).*

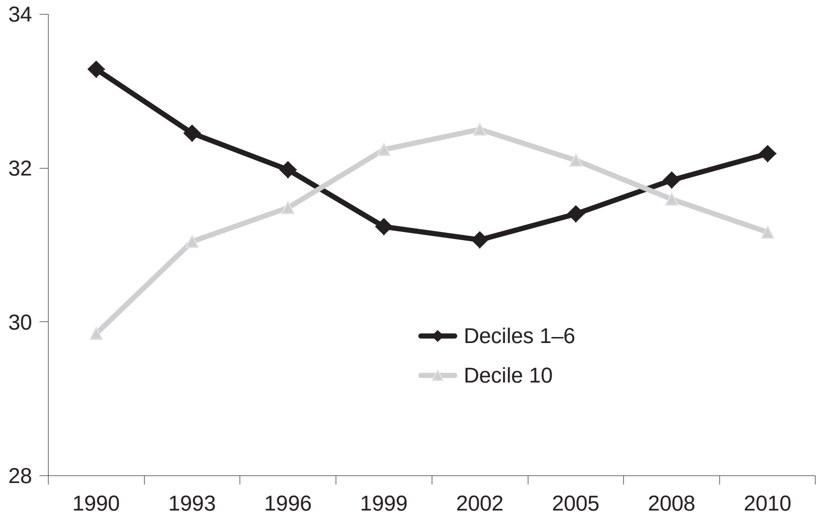


Figure 9.9 Decile shares. Unweighted mean for developing countries, 1990–2010. *Note: The decile shares are computed over the distribution of household consumption per capita. Source: Own estimates based on PovcalNet (2013).*

contrast in the GIC corresponding to the 1990s and the 2000s. The first one is clearly increasing, suggesting growing inequalities, whereas the second is decreasing (and flatter), indicating a fall in well-being disparities in the 2000s. On average, in that decade consumption per capita grew by more than annual 4% in the three bottom deciles of the national distributions and by 3% in the top decile.

Naturally, the contrast between decades is also evident when looking at income shares. The results are summarized in Figure 9.9: whereas the share of the bottom 60% fell 2 points in the 1990s and increased 0.9 points in the 2000s, the performance of the top 10% was almost the exact mirror. The share of the "middle" (deciles 7–9) has remained quite stable over the two last decades (36.9 in 1990, 36.5 in 1999, and 36.6 in 2010). This stratum seems not only quite homogeneous across countries but also over time (Palma, 2011).

9.4.2 Changes by Region

Changes in inequality have been heterogeneous across the six geographical regions of the developing world (Figure 9.10).[45] The mean Gini coefficient in Latin America increased more than two points in the 1990s and then dropped in the 2000s by a larger amount. The data reveals almost no change in inequality in sub-Saharan Africa over the two last

[45] We prefer not to report the regional patterns before 1990 because the number of observations is small in several regions. The numbers of countries by region in the sample we use to assess inequality trends are 8 in EAP, 20 in ECA, 19 in LAC (all in Latin America, none from the Caribbean), 5 in MENA, 4 in SA, and 20 in SSA.

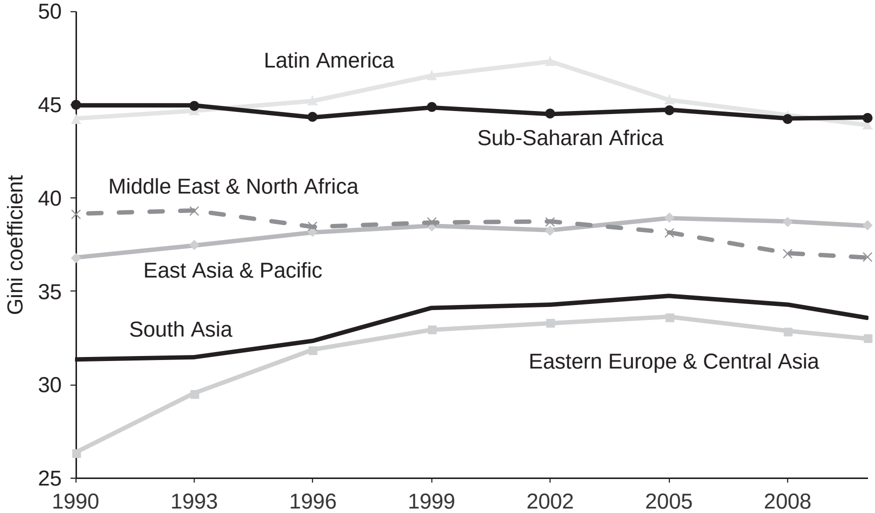


Figure 9.10 Gini coefficients. Unweighted means by region, 1990–2010. *Note: The Gini coefficients are computed over the distribution of household consumption per capita. Source: Own estimates based on PovcalNet (2013).*

decades and some decline in the five MENA countries included in the sample. Instead, the Gini coefficient increased more than two points in Asia and more than six points in Eastern Europe and Central Asia. Figure 9.10 suggests again some pattern toward convergence; the gaps in inequality among regions in the developing world are smaller now than two decades ago. For instance, whereas the gap in the Gini coefficient between Latin America and ECA was 18 points in the early 1990s, it shrank to 11 points in the late 2000s.

In all the regions the share of countries with falling inequality rose in the 2000s, as compared to the 1990s. The two most remarkable changes in the pattern occurred in Latin America and in Eastern Europe and Central Asia. Whereas the Gini went down in 26% of the LA economies in the 1990s, that share increased to 95% in the 2000s. In ECA, whereas the growth in inequality was generalized in the 1990s, more than half of the countries experienced reductions in the 2000s.

Using data from PovcalNet, Chen and Ravallion (2012) report changes in the within component of the global mean-log deviation between 1981 and 2008. This within component is a population-weighted measure of the national inequalities. They find substantial increases in East Asia and Pacific (from 0.125 to 0.256) and Eastern Europe and Central Asia (from 0.128 to 0.225), smaller increases in South Asia (from 0.156 to 0.181), Latin America and the Caribbean (from 0.541 to 0.561), and sub-Saharan Africa (from 0.338 to 0.347) and a fall in MENA (from 0.256 to 0.215). Bastagli et al. (2012) report similar patterns using data from PovcalNet, SEDLAC, and LIS.

The picture of national inequalities in the developing world is similar when using other databases. For instance, the unweighted mean Gini in the *All the Ginis* database assembled by Milanovic grew from 36.2 in 1990 to 40.7 in 1999 and then dropped to 39.7 by 2005. Whereas in the 1990s inequality rose in 63% of the economies in the ATG database, that share dropped to 35% in the 2000s. The recorded increase in the 1990s was generalized across regions, but especially intense in Eastern Europe and Central Asia (9 Gini points), whereas the fall in the 2000s was larger in MENA and Latin America. Cornia and Kiiski (2001), Cornia (2011), and Dhongde and Miao (2013) document similar results using WIID data. We find that the linear (rank) correlation coefficient for the change in the Gini coefficient between 1990 and 2005 recorded in PovcalNet and WIID is 0.776 (0.868), significant at 1%. The corresponding values for the comparison between PovcalNet and ATG are 0.721 and 0.765.

The evidence drawn from the EHII database is also roughly consistent with the patterns discussed earlier. The mean Gini for the developing world remained almost unchanged in the 1980s, increased in the 1990s from 42.5 in 1990 to 47.0 in 1999, and dropped to 46.5 in 2002 (the latest available date).[46] Whereas in 62% of the countries inequality increased in the early 1990s, that share dropped to 55% between 1993 and 1999 and to 49% between 1999 and 2002. The regional patterns are roughly consistent with those described earlier. The main difference is that EHII reveals a dramatic increase in inequality in the Middle East and North Africa (7 Gini points) that is not present in the evidence drawn from household surveys.

In the rest of this section we briefly review the literature on inequality changes in each geographic region of the developing world, while we take a closer look to the story of some particular cases: Brazil, China, India, Indonesia, and South Africa.

9.4.2.1 East Asia and Pacific

The inequality patterns in East Asia and Pacific can be traced based on information from only 8 out of the 24 countries in the region, which nonetheless represent 96% of its total population. This set includes Cambodia, China, Indonesia, Lao PDR, Malaysia, Philippines, Thailand, and Vietnam. There is scattered evidence for Fiji, Micronesia, Mongolia, and Timor-Leste, but information is either lacking or too scarce for American Samoa, Kiribati, Korea, Dem. Rep., Marshall Islands, Myanmar, Palau, Papua New Guinea, Samoa, Solomon Islands, Tonga, Tuvalu, and Vanuatu.

The slightly increasing pattern showed in Figure 9.10 for the unweighted mean of the consumption Gini in EAP hides important differences across countries (ADB, 2012; Chusseau and Hellier, 2012; Ravallion and Chen, 2007; Sharma et al., 2011; Solt, 2009; Zin, 2005). Consumption inequality increased in most economies in the region during the 1990s, with the exception of Thailand and Malaysia. The increase was

[46] These estimates are computed dropping countries with few observations in the period.

particularly strong in China, where the consumption Gini climbed around seven points in that decade. The performance in the 2000s was more heterogeneous; inequality continued increasing in China, Lao PDR, and Indonesia and also went up in Malaysia, but there is evidence pointing to a fall in consumption inequality in Cambodia, Philippines, Thailand, and Vietnam.

Overall, considering the two decades, EAP combines countries with systematic increases in inequality (China, Indonesia, Lao PDR), several cases in which inequality had a cyclical pattern, ending in 2010 at similar levels than in 1990 (Cambodia, Malaysia, Philippines, and Vietnam), and only one successful story of consistent reduction in consumption inequality: Thailand, for which the estimated reduction in the Gini coefficient exceeded five points; from 45.3 to 39.4 over the last two decades. Universal social policies, including basic education and health, have been stressed by many authors as significant drivers of that fall (Jomo and Baudot, 2007).

Probably the most striking phenomenon regarding inequality in EAP was the strong rise that took place in the two most populous countries of the region, China and Indonesia: the Gini coefficient went up around 5 points in Indonesia and more than 10 points in China over the last two decades. Such dynamics happened in a context of high growth and falling poverty, most notably in China. Sharma et al. (2011) summarize the main factors behind these changes: (i) the realignment of activity away from agriculture and toward industry and services; (ii) the skill premium increase due to the unmatched growing demand for skills, and even the emigration of skilled workers; (iii) increasing inequalities in educational attainment in secondary and tertiary schooling; and (iv) a lack of infrastructure linking urban areas with rural areas and other barriers to labor mobility.

China

Since 1978, when China started pro-market reforms, GDP has increased at an average rate close to 10%, and household per capita income has grown more than 7% per year. Such remarkable economic transformation has been accompanied by important changes in inequality and poverty. China is a successful story of reduction in absolute poverty (Minoiu and Reddy, 2008; Ravallion and Chen, 2007, 2008; World Bank, 2009). However, it is argued that it will be harder for China to maintain its past rate of progress against absolute poverty without addressing the problem of rising inequality.

The increase in income inequality in China over the last three decades has been widely documented. Ravallion and Chen (2007) and World Bank (2009) show that income inequality rose from the mid-1980s through 1994, dipping a bit in the late 1990s, and then edging upward thereafter. Li et al. (2013), among others, document the increase in the 2000s, a pattern explained by the widening of the rural–urban income gap, and the increase in income from property and assets, driven by the development of urban residential real-estate markets, the expansion of stock and capital markets, the growth of private enterprises, and other property rights.

A significant share of income inequality in China is now accounted for by rural–urban differences in income levels. The concentration of growth in urban areas is creating a rural–urban divide in employment and earning possibilities, exacerbated by the much more rapid development that occurred in coastal areas. Interestingly, in marked contrast to most developing countries, relative inequality is higher in China's rural areas than in urban areas. However, there has been convergence over time with a steeper increase in inequality in cities.

Indonesia

During the 30 years before the Asian crisis of 1997–1998, which coincided with the New Order under Suharto's dictatorship, Indonesia GDP grew at an average rate of 7% per year. The process was not smooth and went through different phases that implied immense structural change. Despite problems with the data, scholars agree in that there was a systematic drop in poverty rates between 1976 and 1997. At the same time, overall consumption inequality in Indonesia did not change markedly with development until the late 1980s, when inequality started to rise, driven by increasing income disparities in urban areas. Alatas and Bourguignon (2000) decompose the inequality increase between components associated with changes in the structure of earnings, changes in occupational choice, and changes in the sociodemographic structure of the population. They find as main explanations the migration from rural to urban areas and the increase in nonfarm self-employed work. The increase in inequality was partly offset by shrinking income gaps in rural areas (Cameron, 2002). Alatas and Bourguignon (2000) find that the returns to land size decreased between 1980 and 1996; opportunities for off-farm earnings for rural households also contributed to falling rural inequality.

Indonesia was severely hit by a financial crisis: In 1997–1998 GDP dropped by 15%. This turned into a sharp decrease in inequality and an increase in poverty. Skoufias and Suryahadi (2000) find that this pattern seems to have arisen from a decrease in regional inequality. Urban areas (which tend to be wealthier than neighboring rural areas) were hit harder, and the urban middle class, who lost their formal sector jobs, was harshly affected. As the crisis reduced the per capita expenditure of households, the percentage reduction was probably less among the poorer population than among the less poor population. Since 2001, and along with the process of decentralization of powers to local authorities, a general pattern of rising consumption inequality has been observed. Miranti et al. (2013) suggest that the recent increases in inequality may be linked to the higher share of workers employed in the informal sector (70%), hence not covered by minimum wage legislation or employment protections.

9.4.2.2 Eastern Europe and Central Asia

The transition from central planning to market system in the countries of Eastern Europe (EE) and those belonging to the former Soviet Union (FSU) had profound socioeconomic impacts, which could be better documented (compared to the situation in previous decades) due to the improvements in the production and release of information by the new administrations (Milanovic, 1998).[47] The fall of the communist regimes was followed by a substantial increase in inequality in almost all countries.[48] According to PovcalNet data the mean Gini for the distribution of per capita consumption expenditures grew from 26.4 in 1990 to 31.9 in 1996. The increase in the first half of the 1990s was particularly strong in those countries belonging to the FSU and in Southeast Europe and somewhat milder in those economies that joined the European Union. Such developments have been linked to the process of privatization, which implied an increase in earnings dispersion in comparison to the more compressed wage structure of the state-owned firms. One key characteristic of the planned economies was the imposition of wage "grids" that forced a wage compression; the fast transition from wage setting under the wage grids toward a less-regulated labor market provoked a rise in the returns to education, and hence a surge in inequality.[49,50] The economic liberalization also triggered changes in the sectorial structure of the economy; in particular the ensuing de-industrialization during the transition is linked to an increase in inequality (Birdsall and Nellis, 2003; Ferreira, 1999; Ivaschenko, 2002; Milanovic, 1999). Milanovic and Ersado (2010) highlight the role played by the inception or increase of tariffs for utilities, whereas Standing and Vaughan-Whitehead (1995) point to the weakening of the minimum wage as key factors behind the increase in inequality.

After the initial surge in the early 1990s, inequality continued growing in the region in the second half of the 1990s although at slower rates. The patterns were more heterogeneous in the 2000s; inequality increased in some economies, but went down in most countries, especially those in the FSU (World Bank, 2005). The mean Gini for Eastern Europe and Central Asia in the late 2000s was lower than in the late 1990s but still significantly higher (around 7 Gini points) than before the transition.[51]

9.4.2.3 Latin America and the Caribbean

All Latin American countries regularly carry out national household surveys that include income questions, and in some of them also questions on consumption expenditures.[52]

[47] Much information also existed about pre-1990 earnings and household incomes in (former) Czechoslovakia, Hungary, and Poland (Atkinson and Micklewright, 1992).

[48] Milanovic and Ersado (2010); Ivaschenko (2002); Ferreira (1999); Milanovic (1998); Cornia (1996); Cornia and Reddy (2001); Mitra and Yemtsov (2006).

[49] See Fleisher et al. (2005) for a study of 10 transition economies and Gorodnichenko and Sabirianova (2005) for the cases of Russia and Ukraine.

[50] In Hungary, for instance, the income share accrued by the top 1% almost doubled between 1992 and 2009, from 6–7% to 12%. Half of the increase was due to capital income, whereas the other half was due to increased earnings (Mosberger, 2014).

[51] The reader is referred to Chapter 19 for a survey of cross-country studies on the multiple causes of inequality in OECD, including many countries in Eastern Europe.

[52] The increasing availability of surveys in Latin America allowed the creation of databases that make efforts to standardize the generation of poverty and inequality statistics, favoring a close monitoring of the social and labor situation in the region (SEDLAC by CEDLAS and the World Bank, and BADEINSO by UN´s ECLAC).

In contrast, the situation in the Caribbean is much less favorable, as surveys are sporadic, and information is not easy to access. In fact, trends shown in the literature and in this section are restricted to Latin America, which represents 94% of total LAC population.

Latin America experienced two distinct distributive patterns in the last three decades (De Ferranti et al., 2004; Gasparini et al., 2011a,b; IDB, 1999; López Calva and Lustig, 2010). During the 1980s, 1990s, and the crises at the turn of the century, income inequality soared in most countries for which comparable data are available. The mean Gini for the distribution of household per capita income crawled from 50.1 in 1980 to 51.5 in 1986, 51.9 in 1992, 53.0 in 1998, and 53.4 in 2002 (Gasparini et al., 2013). The frequent macroeconomic crises that hit the region in that period were unequalizing because the poor were less able to protect themselves from high and runaway inflation, and adjustments programs frequently hurt the poor and the middle-class disproportionately (Lustig, 1995). The market-oriented reforms that started in Chile in the 1970s and became widespread in the region in the 1990s were associated with rising inequality, although this pattern had a notable exception in the case of Brazil (López Calva and Lustig, 2010). In most countries employment reallocations brought about by trade liberalization and the skilled-biased technical change associated to the modernization of the economy implied a sizeable reduction in the demand for unskilled labor, which led to higher inequality. In some countries adjustments that led to a contraction in the demand for labor affected unskilled workers disproportionately. All these changes took place in a framework of weak labor institutions and safety nets, and hence their consequences made a full impact on the social situation (Gasparini and Lustig, 2011).

Starting in the late 1990s in a few countries and in the early 2000s for the rest, inequality began to decline. The mean Gini for the distribution of household per capita income dropped from 53.4 in 2002 to 50.9 in 2008 (Gasparini et al., 2013). Updated SEDLAC and BADEINSO statistics suggest that the downward trend continued. The evidence, in fact, indicates that between 2002 and 2013 income inequality went down in all Latin American economies. This remarkable decline appears to be driven by a large set of factors, including the improved macroeconomic conditions that fostered employment, the petering out of the unequalizing effects of the reforms in the 1990s, the expansion of coverage in basic education, stronger labor institutions, the recovery of some countries from severe unequalizing crises, and a more progressive allocation of government spending, in particular monetary transfers. The empirical evidence on the driving factors of the recent fall in inequality is, however, still scarce and fragmentary (Cornia, 2011; Gasparini and Lustig, 2011; López Calva and Lustig, 2010).

Brazil

For decades Brazil was singled out as the most unequal economy in Latin America, and in some rankings even the most unequal in the world. In the late 1980s the Gini coefficient for the distribution of household per capita income reached values higher than 60. But

Continued

from that point on inequality started to decrease, first slowly in the 1990s, and then more dramatically in the 2000s.[53] By 2011 the Gini reached an unprecedented low value of 52.7, several points below the level of some other Latin American economies (e.g., Honduras, Colombia, Bolivia).[54] Brazil—the fifth most populous nation in the world—is still a high-inequality country, but it stands out as a successful case of consistent reduction of income disparities.

Data from the Brazil´s national household survey (PNAD) reveals a drop in the Gini of 2 points in the late 1970s and no systematic changes for most of the 1980s, until a deep macroeconomic crisis hit the country, pulling inequality to unprecedented levels. The Gini went up from 59.2 in 1986 to 62.8 in 1989 and returned to 59.9 in 1993. During the 1990s the Gini moved down very slowly, decreasing by just 1 point between 1993 and 2001. That pace drastically increased in the 2000s: the Gini went down from 58.8 in 2001 to 52.7 in 2011, averaging a fall of 0.6 points a year. During 10 years per capita income of the poorest 10% of the Brazilian population grew at an average annual rate of 7%, almost three times the national average.

In an in-depth study of the determinants of inequality changes in Brazil, Barros et al. (2010) highlight the role played by the sharp fall in earnings inequality and the substantial increase in public transfers as the two main direct determinants of the decline in income disparities since the early 2000s.[55] They find that half of the reduction of inequality in labor incomes was associated to the educational progress that took place over the previous decade, which significantly increased the ratio between skilled and unskilled workers. The average years of education for the adult population grew 22% in the 2000s, and the Gini coefficient computed over the distribution of that variable fell 23%, values well above the mean for Latin America (Cruces et al., 2014). Using different decomposition techniques, Barros et al. (2010) and Azevedo et al. (2011) find a sizeable impact of the fall in the returns to education on earnings inequality. Several authors have also found a reduction in spatial and sectorial labor market segmentation. The substantial increase in the minimum wage—68% in real terms between 2002 and 2010—is also underlined as one important force behind the fall in household income inequality, given that the minimum wage sets the floor for both unskilled workers earnings and for social security benefits.

The strong expansion of public transfers accounts to a large share of the fall in income inequality in Brazil (Azevedo et al., 2011; Barros et al., 2010; Alejo et al., 2013; Lustig et al., 2012). The main force was the rapid expansion in the coverage of government cash transfers targeted to the poor, mainly a transfer to the elderly and disabled (*Benefício de Prestação Continuada*) and Brazil's signature conditional cash transfer program *Bolsa Família*.[56]

---

[53] See Barros et al. (2010), Ferreira et al. (2007), Foguel and Azevedo (2007), Hoffmann (2006), and Langoni (2005).

[54] These values are taken from SEDLAC (2013). All sources confirm the strong decreasing pattern in inequality in Brazil.

[55] For evidence and discussion on inequality in the 1990s see Ferreira and Paes de Barros (1999) and Ferreira et al. (2006).

[56] See Fiszbein and Schady (2009) and Veras Soares et al. (2007).

9.4.2.4 Middle East and North Africa

Data constraints are particularly limiting when analyzing distributive issues in Middle East and North Africa. The lack of accessible and comparable household surveys makes it difficult even to identify the extent of poverty and inequality in most MENA countries. The oil-rich economies (Bahrain, Kuwait, Oman, Qatar, United Arab Emirates, and Saudi Arabia) enjoy high levels of per capita income and are usually not included in the analysis of the developing world. In any case, distributive data is rarely available for these economies. A second group, by far the largest in terms of population, consists of middle-income countries. Within this group there is no public accessible information for Lebanon and Libya, just one data point for Djibouti, Iraq, Syrian Arab Republic, West Bank, and Gaza, and only a few for Algeria and Yemen. In sum, the only MENA countries for which it is possible to track changes in poverty and inequality over time are Egypt, Iran, Jordan, Morocco, and Tunisia, but even in these cases data is scattered and often of low quality. MENA has a long way to go to build a reliable, comparable, and sustainable system of household surveys and distributive statistics.

Despite this constraint, several studies shed some light on the trends in inequality in this region.[57] Authors coincide in dividing the last four decades into three periods. The first one, spanning until 1985, was characterized by rapid economic growth. Page (2007) reports a substantial reduction in income inequality between the mid-1970s and the early 1990s.[58] Data from PovcalNet confirms that fall, although the magnitude is more modest. "Middle Eastern economies entered their rapid growth period with income distributions that were becoming more egalitarian, reflecting the political ideology and policies of post-colonial governments" (Page, 2007). The second period covers the late 1980s and most of the 1990s and is characterized by low economic growth and meager or no social gains; real per capita incomes increased by less than 1.5% per year, whereas income distributions were rather stable. The downward pattern in inequality appeared to have resumed in the 2000s, although at a slow pace. According to our estimates based on PovcalNet, the mean Gini fell from 38.7 in 2002 to 36.8 in 2010. These values place MENA as a region of moderate inequality within the developing world, a fact that has puzzled some authors, that would predict higher income disparities given the political process and the balance of political power in those societies.[59]

Alvaredo and Piketty (2014) analyze the issue from a regional perspective and show that, irrespective of the uncertainties on within-country disparities, income inequality is extremely large at the level of the Middle East taken as a whole, simply because regional

[57] See Acar and Dogruel (2012), Adams and Page (2003), Bibi and Nabli (2010), Iqbal (2006), Page (2007), and Salehi-Isfahani (2010).

[58] According to Iqbal (2006) reliable pre-1985 household surveys are only available for Tunisia and Egypt.

[59] "In many MENA countries, from the Maghreb to the Arabian peninsula, power is wielded by rather narrow groups.... Seen from this perspective, the most puzzling thing about inequality in the Middle East is how low it is" (Robinson, 2009).

inequality in per capita GNP is particularly large. Under plausible assumptions, the top 10% income share could be well over 60%, and the top 1% share might exceed 25% (vs. 20% in the United States, 9% in Western Europe, and 18% in South Africa). The authors conclude that the popular discontent that contributed to the Arab spring revolt might reflect the fact that perceptions about inequality and the (un)fairness of the distribution are determined by regional (and/or global) inequality, and not only on national inequality.

9.4.2.5 South Asia

South Asia has been a region of low inequality for developing world standards, though rising since the early 1990s. In India, further discussed in a separate box, the consumption Gini moved from 30.8 in 1993 to 33.9 in 2010. Bangladesh displayed relatively low inequality throughout the 1980s (Gini equal to 26.1 in 1984), but the situation worsened since the beginning of the 1990s: the Gini climbed to 32.1 by 2010. Khan (2008) argues that incomes from nonfarm sources and the high concentration of land tenure have all been disequalizing forces, whereas the positive effects of the more-evenly distributed farm income were offset partly by its declining share in total income.

Scholars do not always agree about the distributive changes in Pakistan; PovcalNet helps defining the picture by providing consumption Ginis of 33.2 for 1990, 28.7 for 1996 and 30.0 for 2008. The high economic growth during the 1980s contributed to a sharp decline in poverty, but it was accompanied by a mild increase in inequality. The fall in economic growth during the 1990s resulted in a rise in poverty, whereas inequality decreased modestly. According to Hussain (2008) in Pakistan there is an institutional structure that excludes a large proportion of the population from the process of economic growth as well as governance.

Sri Lanka experienced rising inequalities between 1985 (Gini of 32.5) and 2007 (Gini 40.3)—among the highest increase in the region during the period of free market reforms, integration to the world markets and high growth—with a reversal of the trend toward 2010 (Gini of 38.3) and persistent regional disparities due to conflict. Nepal presents similar dynamics. Gosh (2012) notes that rising inequality reflects two components: first, growing vertical inequality within the modern industrial sector driven by the returns to skill, and second, increasing disparities between the industrial fast-growing sector and the traditional agricultural activities.

India

Chakravarty (1987) argues that even if policymakers in India adopted a development strategy based on central planning over the 40 years following independence, "there was a tolerance towards income inequality, provided it was not excessive and could be seen to result in a higher rate of growth than would be possible otherwise." One of the explicit goals of the socialist program was to limit the economic power of the elite in the context of a mixed economy. From the mid-1980s, however, India gradually adopted market-oriented economic reforms. Initially, these were accompanied by an expansionist fiscal policy involving allocations to rural areas, to counterbalance the negative redistributive effects of the liberalization. The speed of reforms accelerated during the early 1990s, and the focus shifted away from state intervention toward liberalization, privatization, and globalization.

Most analysis on inequality in India over the last three decades are based on the observations from the expenditure surveys conducted in 1983, 1987/8, 1993/94, 2004/05, and 2009/10 for urban and rural areas, which have allowed for an analysis pre- and postreforms. Inequality increased significantly in the postliberalization years, especially in urban areas; on the contrary, estimates of absolute poverty measures have systematically fallen since 1983. Mazundar (2012) summarize the main drivers of these changes: (i) the lead in employment and output growth has been taken not by manufacturing but by the tertiary sector, which displays higher inequality in pay; (ii) much of the labor reallocated from agriculture is absorbed in the informal sector, where earnings are only slightly higher than the poverty line; (iii) although numerous social insurance schemes have been established, their actual impact has been limited and regressive as they have disproportionately benefited workers in the small formal sector; (iv) the modest and selective increase in social sector spending is constantly threatened by the budget deficit; (v) the education polices implemented over the years have been biased toward the promotion of tertiary education and have neglected basic primary and lower secondary education.

From a different perspective, Banerjee and Piketty (2010) look at the tax-based shares of top incomes. Their results suggest that the gradual liberalization of the Indian economy made it possible for the top 1% to substantially increase their share of total income, from 4.7% in 1980 to 8.9% in 1999. Although in the 1980s the gains were shared by everyone in the top percentile, in the 1990s it was those in the top 0.1% who benefited the most.[60]

9.4.2.6 Sub-Saharan Africa

Although recently there have been many improvements, the lack of a consistent body of household surveys undermines the assessment of income inequality in sub-Saharan Africa. Time series data on inequality is severely lacking in most SSA countries, hindering the inferences about trends in the region. For example, in PovcalNet some

[60] The authors stress that these results could be linked to the debate around the Indian growth paradox of the 1990s. According to the household expenditure survey conducted by the National Sample Survey (NSS), real per capita growth during the 1990s was fairly limited, in sharp contrast with the fast growth measured by national accounts. It was suggested that much of such growth could have gone to the rich, absent from surveys. Banerjee and Piketty (2010) conclude that top incomes could explain between 20% and 40% of the puzzle, which still leaves the bulk of the difference unaccounted for.

SSA countries are missing (Equatorial Guinea, Eritrea, Mauritius, Somalia, Zimbabwe), whereas for 13 of them there is only one observation in the database for the whole period 1981–2010. In fact, very few countries have reliable surveys in the 1980s, and it was not until the mid-1990s when inequality could be really traced with some confidence in the region.

Regional studies typically report a mixed picture, with both increases and decreases in inequality, a fact that could reflect the heterogeneity in the region, but also could be caused by noise in the country estimations (Christiansen et al., 2002; Okojie and Shimeles, 2006). Bigsten and Shimeles (2003), for instance, report that for 17 African countries the trend in inequality shows significant variations over short periods, causing concern about measurement problems.

The available evidence seems to support some few broad facts about consumption inequality in the sub-Saharan African countries. First, inequality is very high on average, possibly the highest in the world. This result is in stark contrast with the presumption of low inequality in SSA, held for a long time based on the predictions of Kuznets-like models and the absence of reliable data.[61] Second, on average inequality does not seem to have changed much in the 1990s and 2000s. Data from PovcalNet and other sources suggest a slow downward pattern; but in any case the evidence is mixed and weak. Third, the heterogeneity among countries in terms of inequality levels and patterns is large, partly possibly due to various measurement errors. It is hard to identify a prototype of an inequality pattern in SSA, as in other regions such as LAC or ECA. The scarce literature on inequality in SSA is consistent with these observations. Go et al. (2007) report that high income inequality levels in SSA have remained more or less constant over the last four decades. Okojie and Shimeles (2006) underline the fact that SSA is one of the most unequal regions in the world, and that disparities have remained persistent over time. In contrast, Sala-i-Martin and Pinkovskiy (2010) picture a more optimistic scenario, reporting a significant downward pattern for inequality during the period of growth (1995–2006).

South Africa

Over the last 30 years there have been important studies of inequality and poverty in South Africa, and a heated debate about trends in post-apartheid transition.[62] South Africa has long been regarded as having one of the most unequal societies in the world. Consistent

[61] Several studies have sought to explain the unexpected result of high inequality in sub-Saharan Africa (Milanovic, 2003; Moradi and Baten, 2005; Okojie and Shimeles, 2006; and Van de Walle, 2008).

[62] See, for example, McGrath (1983), McGrath and Whiteford (1994), Klasen (1997, 2005), Nattrass and Seekings (1997), Terreblanche (2002), Dollery (2003), van der Berg and Louw (2004), Leibbrandt et al. (2009), Leibbrandt et al. (2010a, 2010b), and Aron et al. (2009).

with this view, the country has the highest Gini coefficient of household consumption per capita (63.1 in 2010). During the early 1970s, the previously constant racial shares of income started to change in favor of the blacks, at the expense of the whites, in a context of declining per capita incomes (McGrath, 1983; McGrath and Whiteford, 1994). But while interracial inequality fell throughout the eighties and nineties, inequality within race groups increased (Simkins, 1991; Whiteford and Van Seventer, 2000). Leibbrandt et al. (2010a, 2010b) provide evidence from comparable households' surveys conducted in 1993, 2000, and 2008. These authors find that since the fall of apartheid, inequality continued to increase steadily, both for the whole population and within each racial group. The high level of overall income inequality accentuated between 1993 and 2008, incomes becoming increasingly concentrated in the top decile. van der Berg and Louw (2004) also conclude that rising black per capita incomes over the past three decades have narrowed the interracial income gap, although increasing inequality within the black and Asian/Indian population seems to have prevented any decline in aggregate inequality.

In explaining these changes scholars agree in that the labor market played a dominant role, where a rise in the number of blacks employed in skilled jobs (including civil service and other high-paying government positions) coupled with increasing mean wages for this group of workers. Leibbrandt et al. (2010a, 2010b) indicate that in the initial post-apartheid period participation rates increased faster than absorption rates with a consequent increase in unemployment across all deciles. Since 2000 the aggregate unemployment rate declined, but the in the lower deciles the early post-apartheid trend continued to 2008. State transfers have increased their importance as an income source but not in a way that has substantially narrowed the income gaps. They have, however, compensated for the decreasing share of remittance income.

Increasing inequality and stable poverty are consistent with the rising trend in top income shares recorded between 2002 and 2010 by Alvaredo and Atkinson (2010), which could be associated with the favorable conditions in the world market for agricultural commodities, the increase in the value of minerals other than gold, and the developments in the financial sector.

9.4.3 Inequality Convergence

As suggested earlier, there are signs of inequality convergence among countries in the developing world. As an example, the mean Gini coefficient for the 20 most unequal countries in our PovcalNet sample in 1981 fell 11% in the following three decades, while it increased 58% for the 20 most egalitarian economies. Bénabou (1996) was the first to present empirical evidence for cross-country convergence in income inequality with data from 1970 to 1990 drawn from the Deininger and Squire data set. He found evidence consistent with the predictions of a neoclassical growth model that yields convergence of the entire income distribution and not just the first moment.[63] Evidence on inequality convergence was also found in studies that used improved data: Ravallion (2003) based on PovcalNet, Bleaney and Nishiyama (2003) based on WIID, and Dhongde and Miao (2013) using both data sets. With variations, a typical inequality convergence study estimates

$$G_{it} - G_{i1} = (\alpha + \beta G_{i1})(t-1) + e_{it} \quad \text{for} \quad t = 2, \ldots, T; \quad i = 1, \ldots, N,$$

where G_{it} is the Gini coefficient for country i in year t, and e_{it} is an heteroscedastic error term. The parameter β measures the link between the change and the initial Gini, and therefore $\beta < 0$ indicates inequality convergence. Models could be estimated with the Gini coefficient in levels or logs. In his early study Bénabou (1996) found a β coefficient of -0.039 for a small sample of around 30 countries. Naturally, estimates of β vary according to the data used, the period covered, the time horizons considered, and the regression model applied. Ravallion (2003) estimated a value of -0.028 in the 1990s, Bleany and Nishiyama (2003) a value of -0.0125 between 1965 and 1990, and Dhongde and Miao (2013) a value of -0.022 from 1980 to 2005. This literature has also found that the impact of the initial Gini coefficient on the inequality change diminishes over longer time horizons, and that the speed of inequality convergence is higher than the speed of convergence in per capita income.

We add to this literature our own estimates, taking advantage of the PovcalNet panel of 76 countries from 1981 to 2010 used in this section. Table 9.8 shows the OLS and IVE estimates of α and β for different initial years.[64] The parameter β is negative and significant in all the specifications, suggesting evidence for inequality convergence. The estimated coefficients are in the range of those estimated in the literature.

Although the evidence for inequality convergence in the last decades seems well established, the reasons driving that pattern are not clear. As mentioned before, Bénabou (1996) finds the evidence on convergence consistent within the framework of a growth model. In contrast, the evidence for unconditional inequality convergence is interpreted by Ravallion (2003) as the result of policy and institutional convergence since around 1990, when socialist planned economies became more market-oriented, and non-socialist economies adopted market reforms.

[63] The Benabou model may not be considered strictly neoclassical because it involves market imperfections and endogenous redistributions (see Chapter 14 in this Handbook for an extensive discussion of these issues).

[64] Caselli et al. (1996) and Dhongde and Miao (2013) discuss biases that may arise in an OLS model. However, in the IVE model the instrument used to correct for measurement error may include as much measurement error itself.

Table 9.8 Inequality Convergence
Models of the change in the Gini coefficient

	Initial year 1981			Initial year 1990		
	Intercept (α)	**Slope (β)**	R^2	**Intercept (α)**	**Slope (β)**	R^2
Gini Index						
OLS	1.098 (18.90)***	−0.026 (20.38)***	0.49	0.908 (11.20)***	−0.023 (10.97)***	0.35
IVE	1.271 (17.91)***	−0.031 (17.61)***	0.47	0.855 (9.83)***	−0.021 (9.90)***	0.35
Difference Hausman Test	0.173 (4.26)***	−0.005 (4.03)***		−0.053 (1.69)**	0.002 (2.61)***	
Log Gini index						
OLS	0.118 (28.41)***	−0.032 (28.11)***	0.65	0.105 (15.53)***	−0.029 (15.14)***	0.27
IVE	0.135 (25.79)***	−0.037 (24.75)***	0.63	0.104 (14.07)***	−0.028 (13.66)***	0.27
Difference Hausman Test	0.017 (5.30)***	−0.005 (5.01)***		−0.001 (0.49)	0.001 (0.48)	

Note: Robust *t* statistics in parentheses; **significant at 5%; ***significant at 1%; the heteroskedasticity-consistent covariance matrix estimator is used (HC1). IVE estimates use the initial value as the instrument for the inequality measure in the second survey. The number of observations is 456 in the first panel and 281 in the second.

9.4.4 Absolute Inequality

Although relative inequality has been the preferred concept in empirical work in development economics, absolute views of inequality certainly have some intuitive appeal (Amiel and Cowell, 1999; Atkinson and Brandolini, 2004). Interestingly, the trends in the two concepts over the last decades have been different in the developing world.[65] The fact that most countries experienced economic growth, while at the same time relative inequality did not fall, implied widening absolute income differences. On average, the absolute difference in monthly consumption per capita between the top and bottom 10% of each country increased over the two decades from US$ 415 (PPP adjusted) in 1990, to US$ 497 in 2002, and US$ 646 in 2010. In more than 90% of the countries in the sample that absolute difference was higher in 2010 than in 1990.

The contrast between the recent trends in absolute and relative inequality in the developing countries is illustrated in Figure 9.11. Whereas relative inequality rose in the late 1980s and early 1990s, absolute inequality declined, driven by a reduction in

[65] See, as an example, the analysis of Atkinson and Lugo (2010) for Tanzania.

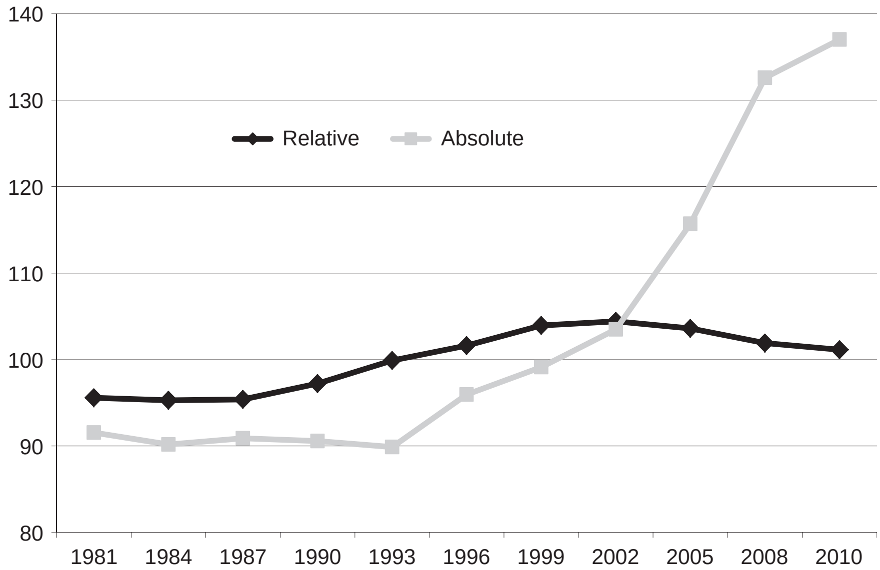


Figure 9.11 Absolute and relative Gini coefficients. Unweighted means, developing countries, 1981–2010. *Note: Normalized to 100 = mean over the period 1981–2010. The Gini coefficients are computed over the distribution of household consumption per capita. Source: Own estimates based on PovcalNet (2013).*

mean income. The strong growth in the developing world since mid-1990s is reflected in the substantial hike in the degree of absolute inequality. Although some equalizing forces operated in the 2000s that reduced the relative gaps, they were not enough to narrow the absolute gaps in a context of economic growth. Based on these facts, Ravallion (2004) argues that the disagreements over whether inequality in the world has gone up or down may partly be due to differing views about the importance of absolute versus relative conceptions of inequality.

9.4.5 Aggregate Welfare

The typical way of assessing the economic performance of a country is by means of its per capita income or output. However, this practice is valid only when the evaluator's welfare function is utilitarian. Except in this extreme case, measuring aggregate welfare involves not only knowing the mean but also other elements of the income distribution, in particular the degree of inequality. Although social welfare functions are naturally arbitrary, because they depend on the analyst's value judgments, it is common in the literature to work with anonymous, Paretian, symmetric, and quasiconcave functions. For

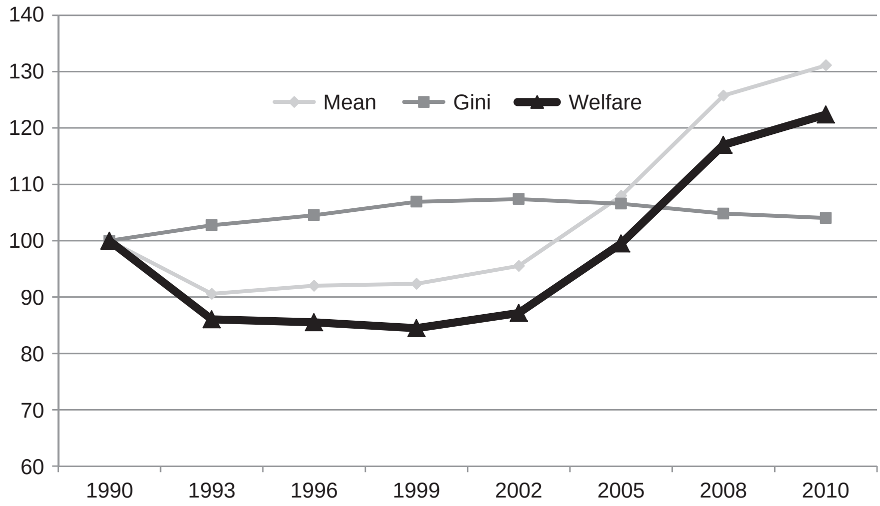


Figure 9.12 Aggregate welfare. Sen welfare function, unweighted mean, developing countries, 1990–2010. *Note: Normalized to 100 = value in 1990. Source: Own estimates based on PovcalNet (2013).*

simplicity, here we consider the abbreviated welfare function proposed by Sen (1976), $W_S = \mu(1 - G)$, where μ is the mean of the distribution and G is the Gini coefficient. Figure 9.12 shows the unweighted mean of W_S for the developing countries in the period 1990–2010 computed from household survey data. In general, aggregate welfare has followed changes in per capita consumption. The fall in mean consumption in the early 1990s (mostly due to the negative performance in ECA) was reinforced by the increase in inequality, driving welfare down by around 15%. Between 1993 and 2002 mean consumption went up, but the change was counterbalanced by a similar increase in the Gini, keeping welfare roughly constant. The 2000s witnessed a robust increase in mean consumption, along with some fall in inequality, implying a 40% increase in aggregate welfare between 2002 and 2010. According to these calculations, the mean aggregate welfare in the developing countries was 22% higher in 2010 than in 1990, implying an annual growth rate of around 1%.

To calculate welfare, it is necessary to have estimates of the mean income and some inequality measure. Ideally, both parameters should be estimated from the same source, typically a household survey, as we have done so far. Some authors have taken a different approach, anchoring the mean to a variable from National Accounts, such as per capita GDP or aggregate household consumption expenditures. For several reasons changes in mean income from household surveys tend to differ significantly from changes in per capita GDP (Anand and Segal, 2008; Deaton, 2003, 2005). Some of these differences are natural because per capita income and GDP are different concepts, but some are

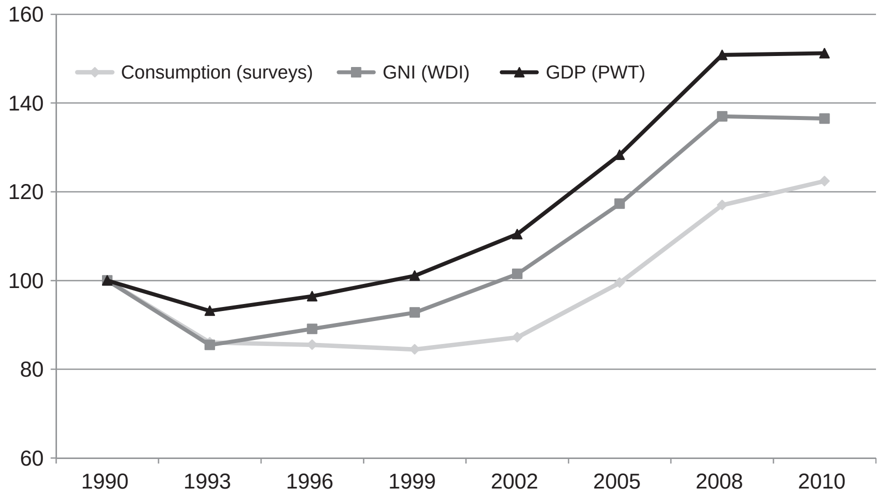


Figure 9.13 Aggregate welfare for alternative mean variables. Sen welfare function, unweighted mean, developing countries, 1990–2010. *Note: Mean anchored to per capita consumption (PovcalNet), GNI per capita (WDI), and GDP per capita (PWT). Normalized to 100 = value in 1990. Source: Own estimates based on PovcalNet (2013), WDI, and PWT.*

rooted in measurement errors both in household surveys and in National Accounts. Some authors pay the price of the potential inconsistency of using two different data sources (i) in order to avoid departing from the typical growth and development literature that is based on National Accounts data, (ii) as a way to alleviate the underreporting issue in household surveys, and (iii) to avoid problems related to the unavailability of surveys for many years in several countries (Ahluwalia et al., 1979; Bhalla, 2002; Bourguignon and Morrison, 2002; Sala-i-Martin, 2006).

In Figure 9.13 we report the results of computing the unweighted average of aggregate welfare across developing countries using alternative mean income variables. According to these estimates, mean welfare in the developing world grew at an annual 1% from 1990 to 2010 using mean consumption per capita from household surveys, 1.6% using per capita GNI from WDI, and 2.1% using per capita GDP from Penn World Tables (PWT; Heston et al., 2012).[66] These discrepancies are worrying and call for increasing efforts to understand and reduce the gaps among data sources.

[66] In fact, this difference comes from the combination of higher growth recorded in the National Accounts in the 1990s compared to household surveys and the opposite result in the 2000s. For instance, although mean per capita GDP slightly fell between 2008 and 2010, mean consumption in household surveys increased at an annual 2%.

The population-weighted mean of the welfare measure grew at a much higher rate, due to the positive performance of several large countries.[67] The growth rate between 1990 and 2010 was 2.3% using household survey data, 3% anchoring mean income to per capita GNI from WDI, and 3.3% when using the Penn World Tables.

9.4.6 Trends from Tax Records

At the moment of writing (2014), the WTID offers estimates of the tax-based shares of top incomes for a small number of developing countries: Argentina, Colombia, India, Malaysia, Mauritius, Uruguay, and South Africa.[68] Ongoing research analyzes the cases of Brazil, Chile, and Ecuador over the last decades. Results for the former colonial territories—being prepared by Atkinson (British colonies), and Alvaredo, Cogneau and Piketty (French colonies)—will be available soon.[69] Consequently, evidence in this respect is still fragmentary, not only because this particular research program is rather recent in what concerns the developing world, but also because of the unavailability of tax data.

The results for the top 1% income share are presented in Figure 9.14, together with survey-based Gini coefficients for six developing countries. Several elements are worth mentioning. First, both sources are not directly comparable for the reasons discussed in Section 9.3. The top share estimates are in general before taxes, while survey Ginis are net of taxes; in addition the units of analysis usually do not match. Second, there is substantial heterogeneity in this group, both in levels and dynamics, compared to the evidence discussed in Chapter 8 for developed countries. It should be borne in mind that the differences in the tax systems across countries imply different income concepts, so that top share levels should be read with this caveat in mind. Third, Leigh (2007), who analyzes 13 developed countries and finds a strong and significant relationship between top income shares and broader inequality measures, concludes that "panel data on top income shares may be a useful substitute for other measures of inequality over periods when alternative income distribution measures are of low quality, or unavailable."

[67] Some authors have computed global welfare, ignoring the division of the world in countries. The evidence suggests an increase in aggregate welfare in the developing world in the last decades (Atkinson and Brandolini, 2010; Pinkovskiy, 2013; Pinkovskiy and Sala-i-Martin, 2009).

[68] The results for China (1986–2003) and Indonesia (1982–2004) available in the WTID are based on household surveys and not on tax records. The fact that top income share estimates are lower than in the most egalitarian developed countries shows that they are likely to be underestimated. In the case of China, the rising trend is robust and can be taken as an indicator of the true dynamics of concentration at the top. Piketty and Qian (2010) show that top income shares increased at a very high rate during the period, which is consistent with the evidence discussed in Section 9.4. The top decile share rose from about 17% in 1986 to almost 28% in 2003—that is, by more than 60%. The top 1% income share more than doubled between 1986 and 2001, from slightly more than 2.6% in 1986 to 5.9% in 2003.

[69] This project has assembled data for some 40 colonies covering the periods before and after independence. Unfortunately, the series stop before 1970 in most cases due to unavailability of recent data.

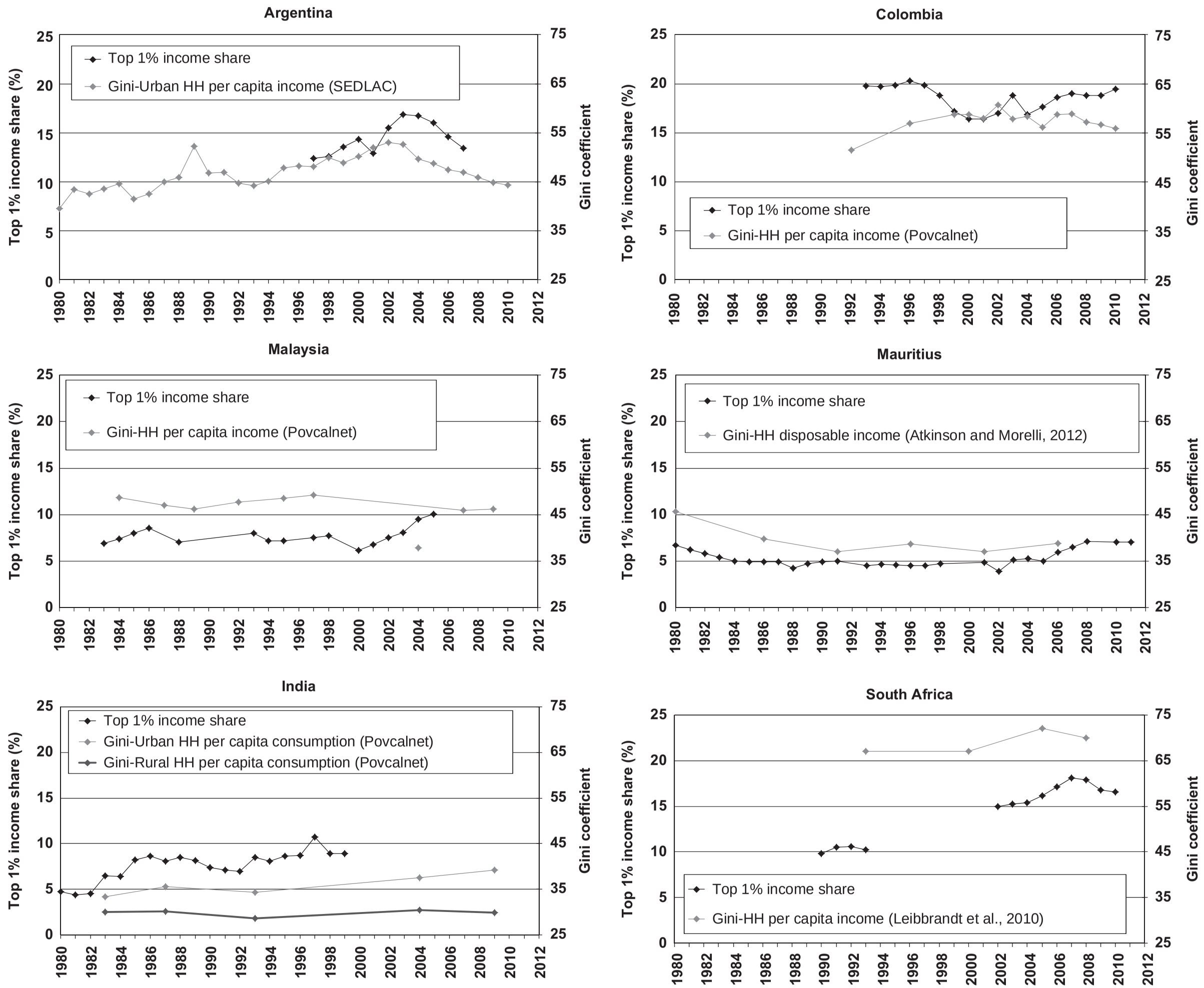


Figure 9.14 Top 1% income share and Gini coefficients. Argentina, Colombia, Malaysia, Mauritius, India, and South Africa. *Note: Gini coefficient for Malaysia 2004 identified as outlier. Source: Own estimates based on WTID and Povcalnet.*

Figure 9.14 seems to agree with the results in Leigh (2007) in some cases but not in others. For example, the trends in the Gini coefficient and the top 1% share in Colombia are particularly diverging, whereas the dynamics in Mauritius are remarkably similar. Top shares and broader, synthetic inequality measures can very well display different trends. The problem arises when the top group plays a major role in the changes in inequality and survey data fail to capture high incomes.

9.4.7 Exploring Inequality Changes

Explaining changes in income distribution is a very difficult and challenging task that lies well beyond the objectives of this chapter. In this section we briefly review some methodologies to study the determinants of the income distribution changes and lay out some of the main results regarding developing countries.[70] Certainly, there has been sustained progress in our understanding of the factors that shape income distributions, but yet the image that emerges from reviewing the literature is still that of a patchwork of numerous hypotheses without conclusive empirical support. In addition, countries may be very heterogeneous in terms of the importance of the various causal factors and their actual effects.

Decompositions are one of the most widely used techniques to characterize income distribution changes. Typically, an income model is estimated, and a counterfactual distribution is simulated modifying some elements of the estimated income model (e.g., parameters or the distribution of observable factors), while keeping the rest fixed. The difference between the actual and the simulated distribution captures the first-round partial-equilibrium effect of the change under study.[71] The method generates entire counterfactual distributions and hence can capture the heterogeneity of impacts throughout the distribution. The decompositions have been typically used to shed light on the impact of changes in the returns to education; in the demographic, sectorial, occupational and educational composition of the population; and in labor and social policies. The decompositions do not allow for the identification of causal effects and suffer from the usual problems of equilibrium-inconsistency and path dependence. Nevertheless, these types of exercises are informative about the relative strength of several direct determinants that may be driving the distributive changes, and therefore could be useful in identifying areas in which to focus the research efforts.

[70] Bourguignon et al. (2008a, 2008b) and Ferreira (2010) are excellent references for methodological issues in recent research on inequality determinants.

[71] See Bourguignon et al. (2005), Barros et al. (2006), and Bourguignon et al. (2008a, 2008b) for methodological proposals; Bourguignon et al. (2005) for applications to Asia and Latin America; and Inchauste et al. (2012) for a recent application to poverty reduction in Bangladesh, Peru, and Thailand. Fortin et al. (2011) and Essama-Nssah (2012) are useful surveys of the economic literature on decompositions.

Ideally, income distribution changes should be studied in a general equilibrium framework because they are the result of complex processes that involve all sorts of effects and interactions throughout the economy. Computable general equilibrium (CGE) models have been applied to study changes in the income distributions around the developing world. These exercises, however, depend critically on parameters and functions that are difficult to estimate and rely on many simplifying assumptions. The more recent macro-micro approach combines a CGE model (the macro component) with a microsimulation (the micro component). CGE models provide a framework to assess consistency of policy alternatives, but lack the necessary disaggregation for the analysis of distributive issues, which is provided by the microsimulations. The macro and micro components of this methodology communicate through aggregate variables such as employment levels and wage rates that are generated by the CGE model and used as inputs in the microsimulations (Bourguignon and Bussolo, 2012; Bourguignon et al., 2008a,b). In a related approach, rather than building a full general equilibrium model of the economy, researchers rely on a reduced-form relationship between a set of observed exogenous variables (such as changes in tariff rates) and a set of sector-level variables (such as industry-skill wage premia), that serve as inputs in the microsimulations.[72]

A very different strand of the literature involves the estimation of cross-country regressions, typically with panel data, where an aggregate measure of overall inequality, such as the Gini coefficient, is linked to various potential causal factors (e.g., Anderson, 2005; Li et al., 1998). Naturally, endogeneity problems are endemic to this approach, which is useful to characterize the structure of correlations among variables, but less successful in identifying causal links.

Most of the literature takes a less ambitious but probably more productive road and focuses on the partial-equilibrium impact of specific shocks and policy changes, using different identification strategies depending on the characteristics of the shock/policy and the data available. Examples of these methodologies include (i) a typical supply and demand approach, where the impact of indicators of trade, technology, or other factors on the relative wage between skilled and unskilled workers is estimated, controlling for relative supply; (ii) the cost function approach, where the impact of several indicators on the share of skilled wages in the total wage bill is estimated, using flexible cost functions (usually a translog cost production function); and (iii) mandated wage regressions.[73] When experiments with random assignment are available, causal links are more clearly identified. For instance, the conditional cash transfer program *Progresa* in Mexico was initially implemented with a random assignment of treated and control rural villages, which allowed a rigorous impact evaluation. Taking advantage of that design, Todd and Wolpin

[72] Ferreira et al. (2010) use this approach to estimate the effect of a trade liberalization episode on the distributions of wages and household incomes in Brazil.

[73] See Anderson (2005).

(2006) estimate a full structural model of behavior, including education, fertility, and labor supply decisions, a model that can be used to simulate the distributive impact of policies and shocks.[74]

The bulk of the distributive analysis in developing countries has focused mainly on the labor market and on public and private transfers, while largely setting aside the role played by other sources of income, such as capital, land rents, and business profits. The neglect of these other factors is essentially due to the fact that household surveys fail to capture these income sources properly.[75] This shortcoming has, for instance, severely limited the study of the impact of the natural resources exploitation on inequality, a relevant topic in several developing countries.[76] Most studies narrow the analysis to particular indicators of the labor market, such as the wage gap between skilled and unskilled labor (the wage premium) or the returns to education. For instance, Bourguignon et al. (2005) show that increases in returns to schooling were large contributors to increasing inequality in East Asia and Latin America in the 1990s.

In what follows we review some general debates on the determinants of recent inequality changes.

Growth and development. As shown in Section 9.3.5, there is a significant negative relationship between inequality and measures of development, such as GNI per capita, in a cross section of countries. From this evidence Ferreira and Ravallion (2009) conclude that "high inequality is a feature of underdevelopment." However, the short- or medium-run relationship between inequality and development has proved to be elusive. There appears to be no evidence in the last decades of a significant correlation between the growth rate of an economy and the change in the inequality level (Dollar and Kraay, 2002; Ferreira and Ravallion, 2009; Ravallion, 2001; Ravallion and Chen, 1997).[77] Ravallion (2007), for example, analyzes 290 episodes in 80 countries in 1980–2000 and finds a correlation coefficient non-significant at the 10% level between the changes in the log of the Gini coefficient and changes in the log of mean income in real terms between successive household surveys. The analysis of more recent data from PovcalNet leads to the same conclusion. Using 473 spells in the period 1981–2010 we find a non-significant coefficient of –0.0094. A similar result applies when restricting the sample to

[74] See also Parker and Skoufias (2001) and Gertler (2004) for impact evaluations studies of the Mexican conditional cash transfer program.

[75] For instance, according to SEDLAC data, on average in Latin America in 2010 the share of labor income in total household income was 82.3%, the share of transfers (including pensions) 13.9%, and the rest of the sources just 3.8%.

[76] See Caselli and Michaels (2013).

[77] A related literature finds no support for a Kuznets curve with longitudinal data (Fields, 2002; Hellier and Lambrecht, 2012).

observations after 1990 or 2000, or when considering longer spells.[78] The data suggests that among both growing and contracting economies, inequality increased about as often as it fell. In the last decades economic growth has been distribution-neutral on average in the developing countries.[79]

Globalization. Much of the recent public and academic debate on inequality changes has been related to the rise in globalization. In the latest decades most developing countries have experienced increasing openness to international trade, capital markets flows, and foreign direct investment. The theoretical channels linking these changes to inequality are multiple and complex, which accounts for the lack of conclusive empirical results (Anderson, 2005; Goldberg and Pavcnik, 2007; Harrison et al., 2011; Rama, 2003; Winters et al., 2004; Wood, 1997). Although studies in cross sections of developing countries are inconclusive in relation to the impact of globalization on inequality, several longitudinal estimates concerning countries taken separately or in small groups reveal a positive correlation between openness and the relative demand for skilled labor (Anderson, 2005; Chusseau and Hellier, 2012; Goldberg and Pavcnik, 2007; Harrison et al., 2011). Trade openness may affect the income distribution through various channels. The traditional Stolper-Samuelson effect predicts a reduction in the skill premium in unskilled-labor-abundant developing countries, a prediction that does not appear to be confirmed by the facts (Feenstra, 2008; Goldberg and Pavcnik, 2007). Although some of the research has pointed then to non-trade factors—such as skill-biased technological change and labor institutions—to explain rising wage gaps, in recent years new mechanisms have been explored through which trade can increase income inequality. These mechanisms include heterogeneous firms and bargaining, trade in tasks, labor frictions, and incomplete contracts (Harrison et al., 2011). In addition, competition among developing countries may increase inequality in middle-income countries (e.g., Latin America) competing with low-income economies.[80] Also, the growing size of the developing world, as new countries enter the world markets, may foster inequality by augmenting the world endowment of unskilled labor. The literature finds that the mechanisms through which globalization affects income distribution are country, time, and case specific, and therefore the impacts of trade liberalization need to be examined in conjunction with other concurrent policy reforms (Goldberg and Pavcnik, 2007). In addition, and due to various limitations, the literature is mostly focused on the static link

[78] The relationship becomes negative, although just slightly significant, when using the change in log real per capita GDP (Penn World Tables) or per capita GNI (WDI) as measures of growth.

[79] Ravallion (2004) argues that, on average, growth is not associated with increases in relative inequality but absolute inequality, and it is these higher absolute gaps between "rich" and "poor" that generate the perception of an unequal growth processes.

[80] The increase in inequality in Latin America has also been explained arguing that it is a region relatively abundant in natural resources, whereas in the onset of liberalization Asian countries were relatively abundant in unskilled labor (Wood, 1997).

between globalization and income distribution that typically operates through changes in relative prices and wages, rather than on the dynamic, more indirect link from trade to growth, and then to poverty and inequality.

Technology and education. Skill-biased technological change has been a popular explanation for the rise in inequality in the developed countries. Changes in technology, such as the use of computers, increase the relative demand for skilled workers driving the skill premium up. This hypothesis is also plausible in the developing world, where globalization increased the transfers of more skill-intensive technologies from the North and fostered imports of capital goods, typically complementary of skilled labor. Several studies find that openness-driven technological transfers tend to increase inequality in emerging countries (Conte and Vivarelli, 2007). The increase in the wage skill premium may be temporary, as the introduction of new technologies requires a transitional period during which skilled workers are employed to adapt the firm to the new technology (Helpman and Trajtenberg, 1998; Pissarides, 1997). The empirical applications usually show evidence on the short- and medium-run effects of the reforms, failing to capture the long-run impact. The generalized fall in inequality in the 2000s in the developing world might be in part attributed to the petering out of the unequalizing initial impact of the liberalizing reforms and technological shocks experienced by many countries in the 1990s.

The increase in education may counteract the effect of skill-biased technological change in the Tinbergen´s race between education and technology.[81] In fact, education has expanded in the developing world at high rates during the last decades, mitigating the impact of other factors that tend to increase the wage premium.[82] However, the link between education and income inequality may not be that straightforward. Given the convexities in the returns to education, even an equalizing increase in schooling may generate an unequalizing change in the distribution of earnings. Bourguignon et al. (2005) have labeled this phenomenon "the paradox of progress," a situation in which an educational expansion is associated with higher inequality.[83]

Market reforms. Several developing countries have implemented market-oriented reforms in the last decades, reducing regulations and privatizing firms. The paradigmatic case includes the former socialist planned economies in ECA, but the transition from centrally planned to market-oriented economies was also experienced by several African and Asian countries, including China. The evidence suggests a significant increase in inequality over the transition period. That surge has been linked to the process of privatization, that implied an increase in the earnings dispersion in comparison to the more compressed

[81] According to Tinbergen (1975) changes in earnings inequality are the outcome of a "race" between technological progress raising the demand for skills, and the expansion of education raising the supply of skills.

[82] See Gasparini et al. (2011a, 2011b) for Latin America.

[83] Inequality may also increase after an education expansion, given that wage dispersion is larger at higher educational levels (Alejo, 2012).

wage structure of the state-own firms, and the institutional and regulatory reforms that have increased competition in product and factor markets and decreased the bargaining power of labor.[84] Other non-socialist economies also adopted market-friendly reforms; Ravallion (2003) argues that in some cases (e.g., Brazil) pre-reform controls benefited the rich and kept inequality high, and then reforms help lowering inequality, while in some others (e.g., India) the controls (and the reforms) had the opposite effect.

Fiscal and social policy. Developing countries are characterized by relatively low levels of taxation, heavy reliance on regressive revenue instruments, and low coverage and benefit levels of transfer programs (World Bank, 2006a,b). This structure limits the redistributive potential of fiscal policy and in some cases even exacerbates the market income disparities.[85] Although average tax ratios for advanced economies exceed 30% of GDP, ratios in developing economies (excluding emerging Europe) generally fall in the range of 15–20% of GDP (Bastagli et al., 2012). Tax collection is not only lower but also more regressive than in developed countries. The difficulties in collecting more progressive taxes are related to the high levels of self-employment and sizeable informal sectors, which limit the capacity of the tax authorities to verified taxpayers´ income and assets. On the spending side, in most developing economies social spending is relatively low, and participation in social insurance schemes is restricted to high-income workers in the formal sector and to public-sector employees.[86] All these factors combine for a low redistributive impact of the fiscal policy. For instance, Goñi et al. (2008) and Lustig (2012) finds that the tax and transfer system in Latin America decreased the market Gini by only 2 points, a meager impact compared to the 20-points impact estimated in 15 European economies.

Since the mid-1990s there have been some encouraging signs of improvement, especially in terms of increasing coverage and better targeting of social policies. The recent expansion of conditional cash transfer programs (CCTs) implies a promising approach for enhancing the distributive impact of public spending in developing economies. CCTs typically transfer income to poor households, conditional on households making certain investments on their children's human capital—education, health, and nutrition. Such programs have been adopted in many developing economies, including some sub-Saharan African countries, although on a smaller scale (Fiszbein and Schady, 2009; Garcia and Moore, 2012). CCTs became particularly popular in LAC: by 2010 there were 18 countries in Latin America and the Caribbean applying CCTs, covering

[84] See Cornia (1996), Milanovic (1998), Ferreira (1999), Cornia and Reddy (2001), and Milanovic and Ersado (2010).

[85] For instance, Lustig (2012) finds that in some Latin American countries when indirect taxes are taken into account, the net income of the poor and the near poor can be lower than before taxes and cash transfers.

[86] ILO (2010) reports that in the early 2000s the share of the population above the legal retirement age receiving a pension in developing economies was, on average, around 40%, as compared to 90% in European economies.

20% of total LAC population, and spending on average 0.40% of GDP (Cruces and Gasparini, 2012). Soares et al. (2009) estimated that the CCTs in Brazil and Mexico reduced the Gini for disposable income by 2.7 points, accounting for about a fifth of the decrease in that index between the mid-1990s and the mid-2000s.

Macroeconomic crises. The scale of the recent crisis has placed the distributive impact of macroeconomic shocks back on the agenda. Banking crises, crashes in stock and real estate markets, and GDP collapses are events with potential large effects on the income distribution. Atkinson and Morelli (2011) is the first paper in addressing this issue from an empirical, historical and global perspective. They investigate the effect of crises on inequality as well as the impact of inequality on the probability of economic crises, by analyzing the history of banking, consumption, and GDP collapses over a 100-year period in 25 countries, out of which only 6 are developing economies. These authors observe the variation in distributive variables taking a 5-year window before and after the crisis date, and classify each one according to whether inequality was increasing, constant, or decreasing before and afterward.[87] Table 9.9, panel A, reproduces their results specifically regarding GDP collapses.[88] They identify 103 crises, but for only one-third is information available on inequality changes. The shadowed diagonal shows combinations where the trajectory was unchanged; above the diagonal are cases where the trajectory "bent" downward; below the diagonal are cases where the trajectory "bent" upward. As it is readily apparent, one cannot draw firm conclusions: (i) the raw totals show that most crisis did not involve changes in inequality ex-post; (ii) the number of cases above the diagonal is low and not very different from the cases below the diagonal, which means that GDP crises are not necessarily associated with a specific direction in the change of inequality; and (iii) the inverted V shape (inequality increasing and then decreasing) is not prevalent. Atkinson and Morelli (2011) conclude that "economic crises differ a great deal in whether or not they were preceded by rising inequality, and, in any case, where there was such a rise, causality is not easy to establish." When banking crisis are analyzed instead of GDP drops, the cases in which inequality tend to increase following a crisis are in the majority.

We replicate their methodology for the years 1980–2010 to take into account the set of developing countries and show the results in panel B of Table 9.9.[89] Even if our list is not exhaustive and could be considerably improved, we identify at least 67 crises

[87] In the case of the Gini coefficient, a change is considered significant when it is higher than 0.7% points (that is, 1/3 of 2 Gini points).

[88] A collapse is identified as a cumulative percentage drop in per capita GDP (from peak to trough) of at least 9.5% for 1911 to 1950 and of 5% for the post-1950 period. Their results are somewhat different when they look at banking crises and consumption collapses, but not more conclusive.

[89] Given the higher volatility of per capita GDP in developing countries, we have kept the threshold of a 9.5% drop to identify a crisis for 1980–2010. The data come from the World Bank Development Indicators. Changes in Gini coefficient are taken from PovcalNet.

Table 9.9 Inequality and GDP collapses

Panel A (Atkinson and Morelli, 2011)

		After crisis				
		Decreasing	No change	Increasing	Unknown/ excluded	Totals
Before crisis	Increasing	1	3	2	3	9
	No change	1	4	1	9	15
	Decreasing	1	1	0	1	3
	Unknown/ excluded	1	8	3	64	76
	Totals	4	16	6	77	103

Panel B (developing countries, 1980–2010)

		After crisis				
		Decreasing	No change	Increasing	Unknown/ excluded	Totals
Before crisis	Increasing	0	2	1	0	3
	No change	0	0	1	0	1
	Decreasing	0	1	0	0	1
	Unknown/ excluded	3	3	10	46	62
	Totals	3	6	12	46	67

Source: Own calculations based on PovcalNet (2013) and WDI.

episodes. As in general they occurred during the 1980s or early 1990s, it is not surprising that in most cases inequality changes before the crises remain unknown due to data unavailability. There is a tendency for inequality to rise after a GDP collapse (10 cases), but again the numbers are too small to draw conclusions, and this could just be the continuation of a previous tendency. This is not necessarily in contradiction with Atkinson and Morelli (2011) due to at least two reasons: (i) GDP crises may well be more correlated with financial crises in the developing world, and (ii) such conclusion is highly influenced by the experience of the transition economies after the fall of the Berlin Wall. It should also be noted that several of the canonical Latin American exchange rate crises of the 1980s and 1990s, with the exception of Argentina and Brazil (included in Atkinson and Morelli, 2011), do not fall within our classification of a collapse. In this sense, there is much work to be done about the magnitude of a crisis and its sensitivity on the two-way relationship with inequality. In any case, the pattern in Latin America points to an increase in inequality before the crashes (regressive inflation tax, rise in unemployment due to openness to trade, and loss of competitiveness from exchange rate mismanagement), then followed by short-term reductions after stabilization programs.[90]

Others. Of course this brief review does not exhaust the multiple factors behind distributive changes in the developing world; in fact, arguably any shock or policy could affect the income distribution. For instance, demographic factors, such as the decline in fertility, the rise in life expectancy, and the growing importance of assortative mating and single-parent households have been identified as relevant sources of inequality changes. Labor policies are a key target for research, as well. Several studies find that the weakening of labor institutions such as unions and the declining real value of minimum wages were responsible for the increase in earnings inequality in several developing countries, especially in the 1990s, whereas more ambitious labor policies contributed to the reduction in inequality in the 2000s. Migration and sector changes are also determinants of inequality changes, studied at least since the seminal contributions by Lewis (1954) and Kuznets (1955). Changes in inequality are associated with the geographic and sectorial pattern of growth (Loayza and Raddatz, 2010). Ferreira and Ravallion (2009), for instance, report that in Indonesia a large share of the increase in inequality was associated with migration from wage employment in agriculture to urban self-employment.[91]

9.5. POVERTY: LEVELS

The vast literature on poverty measurement suggests that there are neither normative nor objective arguments to set an unambiguous threshold below which everybody is poor and above which everyone is nonpoor (Deaton, 1997). Despite this central conceptual

[90] The case studies are numerous; see, for example, Forbes (2011) and Lustig (1990).
[91] The role of migration has been particularly relevant regarding global inequality (Milanovic, 2012).

ambiguity, reducing poverty is a deliberate policy objective for governments around the world. The international community has embraced this goal as reflected in the first Millennium Development Goal of halving poverty by 2015. In this section we focus on measures of poverty in the income/consumption space using international poverty lines in terms of U.S. dollars adjusted for purchasing power parity (PPP). This choice implies taking a one-dimensional, monetary, static, absolute view of poverty, that certainly has many limitations and drawbacks, but it is still the best available paradigm to summarize deprivations in the world.

The $1-a-day per person at PPP is a poverty standard meant to define an international norm to gauge the inability to pay for food needs. The $1 line, proposed in Ravallion et al. (1991) and used in World Bank (1990), was chosen as being representative of the national poverty lines found among low-income countries. The line was recalculated in 1993 PPP terms at $1.0763 a day (Chen and Ravallion, 2001), and more recently in 2005 PPP at $1.25 a day (Ravallion et al., 2009).

To make international comparisons of economic aggregates researchers have long favored the use of PPP conversion rates, instead of market exchange rates, with the aim of ensuring parity in terms of purchasing power over both internationally traded and non-traded goods. The simplicity of the PPP adjustment, however, entails several potential drawbacks for the international comparison of poverty measures, that have been extensively discussed in the literature.[92] For instance, a concern is that the weights attached to different commodities in the conventional PPP rate may not be appropriate for the poor.[93] The main data sources for estimating PPPs are the price surveys carried out within countries for the International Comparison Program (ICP) (World Bank, 2008). Although the estimation of PPP conversion rates has substantially improved in the last decades, it still has many limitations, including high variability; the switch from 1993 to 2005 PPP figures led to significant changes in absolute poverty measures in some countries.[94]

The $1.25 line is usually deemed too low for middle-income countries; for that reason it is typical to compute poverty with the $2-a-day standard, which is close to the median of the official poverty lines chosen by developing countries. Although these international lines have been criticized, their simplicity and the lack of reasonable and easy-to-implement alternatives have made them the standard for international poverty comparisons.[95] Although the measurement of poverty with national lines takes into consideration that societies differ in the criteria used to identify the poor, the international

[92] See Anand and Segal (2008), Deaton and Dupriez (2008), Deaton and Heston (2010), Ravallion (2010), and Chen and Ravallion (2001).

[93] Deaton and Dupriez (2008) estimate PPPs for the poor for a subset of countries; the results do not suggest that the reweighting has much impact on the conversion rates.

[94] Likely, the next round of PPP indicators (2011) will be affected by the same type of problems.

[95] For the debate on the international measurement of poverty, see Reddy and Pogge (2010), Deaton (2010), and Chapter 11 in this volume. Gentilini and Sumner (2012) compute global poverty using the national poverty lines officially set in each country instead of using international poverty standards.

lines are unavoidable instruments to compare absolute poverty levels and trends across countries and provide regional and world poverty counts.

The World Bank is the main institution that regularly produces information on poverty measurement in the developing world drawn from original microdata from household surveys.[96] In 2013, the World Bank released an update of the developing world's poverty estimates for 1981–2010. The new poverty estimates combine the PPP exchange rates for household consumption from the 2005 International Comparison Program with data from more than 850 household surveys across 127 developing countries. In this section we rely heavily on that data set (PovcalNet).

The problem of the choice of the welfare variable discussed for inequality in Section 9.3 applies to the measurement of poverty, as well. Although poverty estimates in PovcalNet refer to consumption deprivation, in most countries in Latin America and a few others in the rest of the world, they are constructed from income data. After computing consumption and income poverty in 22 household surveys of 7 Latin American countries using the \$2 standard, we find that on average the ratio of consumption/income poverty is 0.97 with only small differences across countries. Given this piece of evidence we decided not to perform an adjustment for income poverty figures in the analysis that follows.

9.5.1 Income Poverty in the Developing World

Although poverty is a ubiquitous characteristic of the developing economies, its severity widely varies across countries. Figure 9.15 shows the poverty headcount ratio in most of the developing countries in the world, using the \$2-a-day poverty line. The figure reveals the enormous differences among developing nations in terms of monetary deprivation. Although there are economies where the proportion of the population living with less than \$2 a day is below 2%, in several countries that proportion exceeds 80%. The problem of absolute income poverty has a radically different scale in some countries compared to others, even within the developing world.

In 2010, 41% of the population in the developing world lived with less than \$2 a day. The unweighted mean headcount ratio was significantly lower: In a typical developing country 33% of the population was poor according to that criterion. The difference between the weighted and unweighted mean is not determined by China, as the incidence of poverty in that country is similar to the developing world mean, but by India (and to a lesser extent Indonesia and Pakistan), where the deprivation measures are substantially higher. In fact, when ignoring India both the weighted and unweighted headcount ratios become very close (33.3 and 32.7). The median poverty rate is also lower than the mean (23.5 for the \$2 line). Table 9.10 reports these results for other indices and poverty lines. Interestingly, when using the \$1.25 line the weighted mean is

[96] Ahluwalia et al. (1979) was an early attempt to measure poverty in the developing countries.

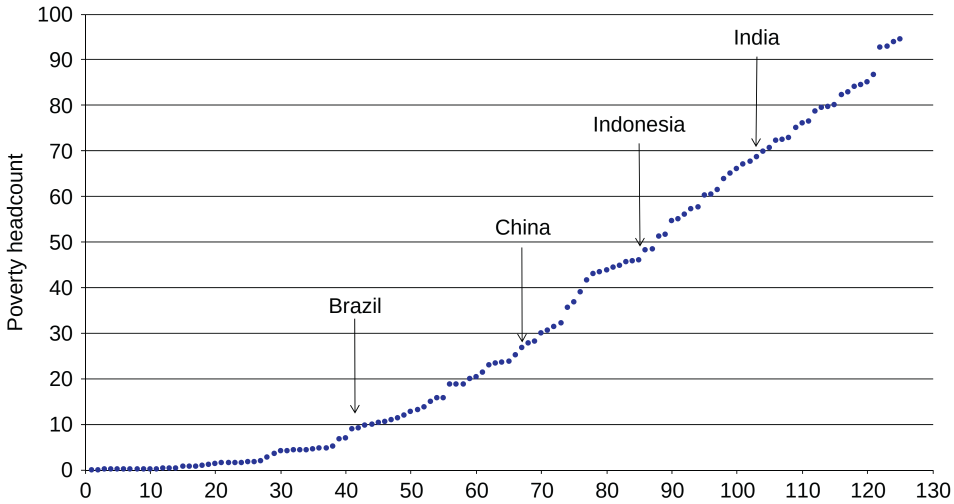


Figure 9.15 Poverty headcount ratio. Developing countries, 2010. *Note: Poverty computed over the distribution of consumption/income per capita with the PPP-adjusted \$2-a-day line. Source: Own calculations based on PovcalNet (2013).*

Table 9.10 Poverty measures
Developing countries, 2010

	Weighted mean	Unweighted Mean	Unweighted Std. dev	Unweighted Median
\$1.25 poverty line				
Headcount	21.4	19.6	22.9	9.3
Poverty gap	6.5	7.5	10.7	2.9
Squared poverty gap	3.0	4.0	6.5	1.0
\$2 poverty line				
Headcount	41.0	33.0	30.1	23.5
Poverty gap	15.8	14.5	16.4	7.1
Squared poverty gap	8.1	8.3	10.9	4.0
\$4 poverty line				
Headcount	66.7	54.8	33.6	59.3
Poverty gap	35.9	29.6	24.0	25.4
Squared poverty gap	22.9	19.5	18.5	13.2

Note: Poverty computed over the distribution of consumption/income per capita.
Source: Own calculations based on PovcalNet (2013).

lower than the unweighted mean for the poverty gap and the squared poverty gap, a result driven by the relatively low value of these indicators in China and Indonesia.

The picture of poverty in the developing world is not significantly affected by changing the poverty indicator or the poverty line. The correlations across countries when using alternatively the headcount (H), the poverty gap (PG) and the squared poverty gap (SPG) with a given poverty line are all higher than 0.9.[97] For a fixed indicator the correlations are higher than 0.95 when changing the poverty line. The correlations are only slightly lower when changing both the indicator and the line (e.g., 0.85 for SPG with the $1.25 line and H with the $2 line).

The top 10 steps in the poverty ladder, using the headcount ratio with the $2 line, are all occupied by sub-Saharan African countries.[98] The following 10 features also eight SSA economies, in addition to a Caribbean country (Haiti) and a South Asian nation (Bangladesh). However, given its size, India is the country with the largest number of poor people. Although around 840 million people in that country live with less than $2 a day, the number in the second nation in that ranking, China, is less than a half (359). Both countries are home of 52% of the poor in the world, whereas the following four countries—Nigeria, Bangladesh, Indonesia, and Pakistan—represent 19%. Of course, these exact figures are valid only for a specific definition of income poverty, but the main results are robust to changes in indices and poverty lines.[99]

As expected, the relationship between mean consumption and poverty is very tight (Figure 9.16, panel a). A simple model of the headcount ratio ($2 line) on log mean consumption per capita estimated in a cross section of developing countries for 2010 accounts for more than 70% of the variation in the data. The cross-country relationship between poverty and inequality is much looser (panel b) see previous comment. The correlation coefficient between the headcount ($2 line) and the Gini is 0.17 (barely significant at 5%). The relationship is somewhat tighter with other poverty indices, but still in all cases the correlation coefficients are lower than 0.3. The magnitude of the correlations is similar when considering different income shares as measures of inequality.

Table 9.11 shows some simple regressions aimed at characterizing the relationship between poverty, mean income, and inequality in a cross section of developing countries. The results of course do not have any causal implication, and they are not helpful to orient policy, but nonetheless are illustrative of the empirical relationship among these three variables. An increase (cross-country) of 1% in mean consumption is associated to a fall

[97] The results are also robust to other poverty measures that do not belong to the FGT family, such as the Sen and Watts indices.

[98] Rwanda, Mozambique, Malawi, Nigeria, Tanzania, Zambia, Madagascar, Burundi, Congo Dem. Rep., and Liberia.

[99] The share of India and China in the world poor is 47% with the $1.25 line, 52% with the $2 line and 51% with the $4 line.

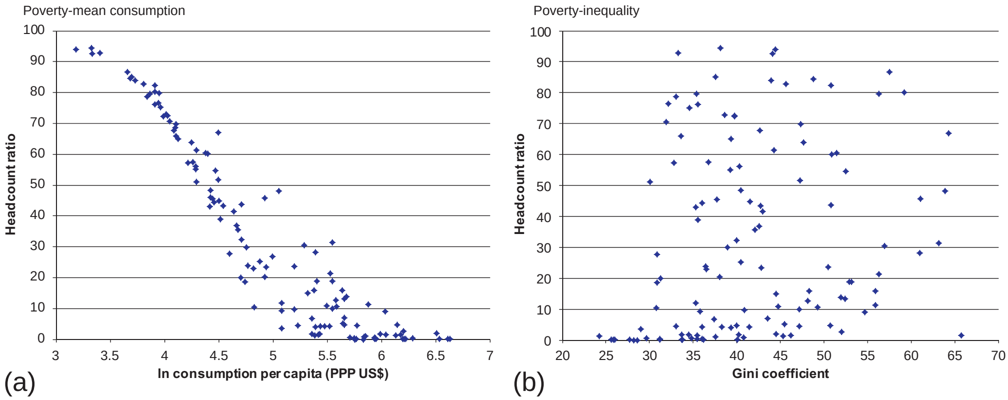


Figure 9.16 Poverty, mean consumption, and inequality. Developing countries, 2010. *Note: Poverty computed over the distribution of consumption/income per capita with the PPP-adjusted 2\$-a-day line. Source: Own calculations based on PovcalNet (2013).*

of around 2% in the headcount ratio, whereas a drop of 1% in the Gini coefficient is associated to a reduction of around 3.3% in poverty measured by the headcount. The results are similar when measuring deprivation with the squared poverty gap.

9.5.2 Poverty by Region

Poverty has a clear regional component: Table 9.12 reveals that Eastern Europe and Central Asia is always the region with the lowest income poverty, followed by Middle East and North Africa and Latin America and the Caribbean. Poverty in South Asia is substantially larger than in Eastern Asia when weighting by population, but roughly similar when ignoring weights. All income poverty measures are substantially higher in sub-Saharan Africa than in the rest of the developing world.

Figure 9.17 unveils the considerable heterogeneity within each geographic region. When using the \$2 line, the poverty headcount ratio ranges in EAP from 1.4 (Malaysia) to 70.6 (Timor-Leste), in ECA from 0.1 (Slovenia) to 35.6 (Georgia), in LAC from 1.2 (Uruguay) to 80.1 (Haiti), in MENA from 1.6 (Jordan) to 45.6 (Yemen), in SA from 6.8 (Maldives) to 76.5 (Bangladesh), and in SSA from 1.5 (Seychelles) to 94.5 (Liberia). Figure A.1 in the appendix displays a map of the poverty levels in the world that illustrates the regional differences, as well as the within-region heterogeneities.

There is a considerable degree of spatial correlation of poverty measures across countries. The Moran scatterplot is a way to illustrate that spatial correlation (Figure 9.18). The horizontal axis shows the normalized headcount ratio of a country (\$2 line), whereas the vertical axis depicts a weighted average of its neighbors´ normalized poverty rates, where neighborhood is defined in terms of geographical proximity. The graph suggests a strong positive correlation between a country poverty incidence rate and that of its neighbors (the Moran correlation coefficient is 0.507, significant at 1%). Almost 80%

Table 9.11 Regressions of poverty measures
Developing countries, 2010

	Log headcount ratio				Log squared poverty gap			
	(i)	**(ii)**	**(iii)**	**(iv)**	**(v)**	**(vi)**	**(vii)**	**(viii)**
ln γ	−2.0 (0.114)★★★	−2.0 (0.085)★★★	−2.1 (0.086)★★★	−2.0 (0.086)★★★	−2.1 (0.127)★★★	−2.0 (0.092)★★★	−2.2 (0.084)★★★	−2.1 (0.091)★★★
ln Gini		3.32 (0.329)★★★				3.76 (0.349)★★★		
Share d1			−0.74 (0.073)★★★				−0.91 (0.071)★★★	
Share d1–d6				−0.12 (0.011)★★★				−0.13 (0.012)★★★
Constant	12.6 (0.577)★★★	0.1 (1.30)	14.8 (0.484)★★★	16.1 (0.561)★★★	11.0 (0.64)★★★	−3.1 (1.39)	13.8 (0.473)★★★	15.1 (0.586)★★★
R^2	0.72	0.85	0.85	0.85	0.68	0.84	0.87	0.85

Note: Poverty computed over the distribution of consumption/income per capita with the PPP-adjusted $2-a-day line. ln γ = log mean household consumption/income per capita; share d1 = share of decile 1 in the household consumption/income per capita distribution; share d1–d6 = cumulative share of deciles 1–6. Robust standard deviations are shown under the coefficients. ★★★=significant at 1%.
Source: Own calculations based on PovcalNet (2013).

Table 9.12 Poverty indicators by region
Developing countries, 2010

	East Asia & Pacific	Eastern Europe & Central Asia	Latin America & Caribbean	Middle East & North Africa	South Asia	Sub-Saharan Africa
Weighted						
$1.25 line						
Headcount	12.5	1.0	5.5	2.4	31.0	48.5
Poverty gap	2.8	0.3	2.9	0.6	7.1	20.9
Squared poverty gap	0.9	0.2	2.1	0.2	2.4	11.8
$2 line						
Headcount	29.7	2.4	10.4	12.0	66.7	69.9
Poverty gap	9.7	0.7	4.7	2.8	23.4	35.7
Squared poverty gap	4.2	0.3	3.2	1.0	10.5	22.4
Unweighted						
$1.25 line						
Headcount	17.9	1.6	8.6	3.5	17.6	41.5
Poverty gap	5.1	0.5	4.0	0.9	3.9	16.6
Squared poverty gap	2.3	0.2	2.6	0.3	1.3	8.8
$2 line						
Headcount	38.4	5.2	15.6	13.5	43.5	62.2
Poverty gap	13.8	1.5	7.0	3.5	14.1	30.2
Squared poverty gap	6.7	0.6	4.4	1.4	6.1	18.1

Note: Poverty computed over the distribution of consumption/income per capita.
Source: Own calculations based on PovcalNet (2013).

of the countries are either in the HH cells (high poverty for the country and its neighbors) or in the LL cells.

The poverty gap indicator has an intuitive-appealing interpretation: when normalized by the poverty line and the total population of a country, it gives the total cost needed to end poverty, in the particular case in which cash transfers could be perfectly targeted to poor people in the amount just needed to reach the poverty line, and no changes in behavior take place. Table 9.13 shows the unweighted mean across countries of the cost of eliminating poverty as percentage of GDP under this scenario in each region. Although the context is clearly unrealistic, the figures give a rough idea of the magnitude of the task of fighting poverty in each region of the developing world in relation to the available economic resources. Although eliminating poverty with the $2 line in this scenario would require on average less than 1 GDP point in the economies of ECA and

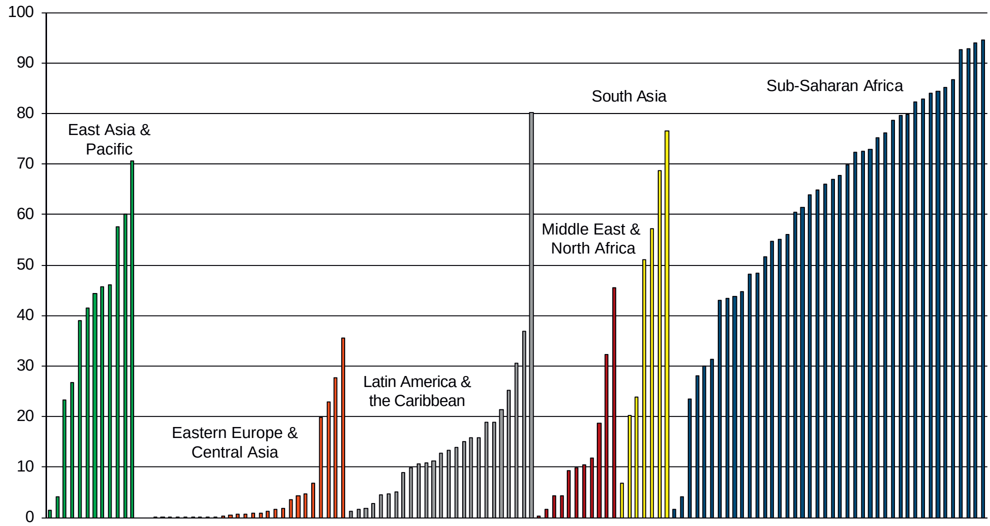


Figure 9.17 Poverty headcount ratio. Developing countries, 2010. *Note: Poverty computed over the distribution of consumption/income per capita with the PPP-adjusted $2-a-day line. Source: Own calculations based on PovcalNet (2013).*

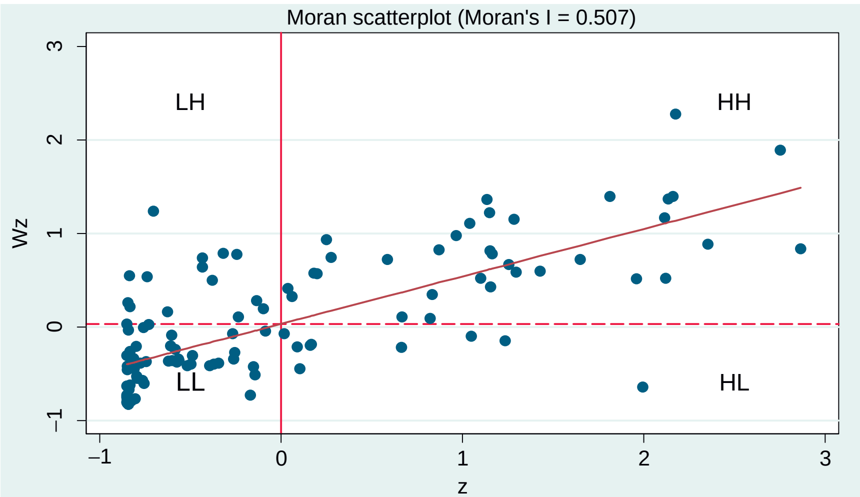


Figure 9.18 Spatial correlation of poverty rates. Moran´s scatterplot. Developing countries, 2010. *Note: Poverty computed over the distribution of consumption/income per capita with the PPP-adjusted $2-a-day line. z is the normalized poverty headcount ratio (the value minus the mean, divided by the standard deviation), Wz is the weighted average of the normalized poverty headcount ratios of a country´s neighbors, where the weights W are defined in terms of contiguity. Source: Own calculations based on PovcalNet (2013).*

Table 9.13 The cost of eliminating poverty
Total poverty gap as percentage of GDP
Unweighted means by region, 2010

	Poverty lines	
Region	**$1.25**	**$2**
East Asia and Pacific	0.6	2.9
Eastern Europe and Central Asia	0.1	0.3
Latin America and the Caribbean	0.7	1.7
Middle East and North Africa	0.2	1.0
South Asia	1.2	6.6
Sub-Saharan Africa	8.1	23.0

Note: Poverty computed over the distribution of consumption/income per capita.
Source: Own calculations based on PovcalNet (2013).

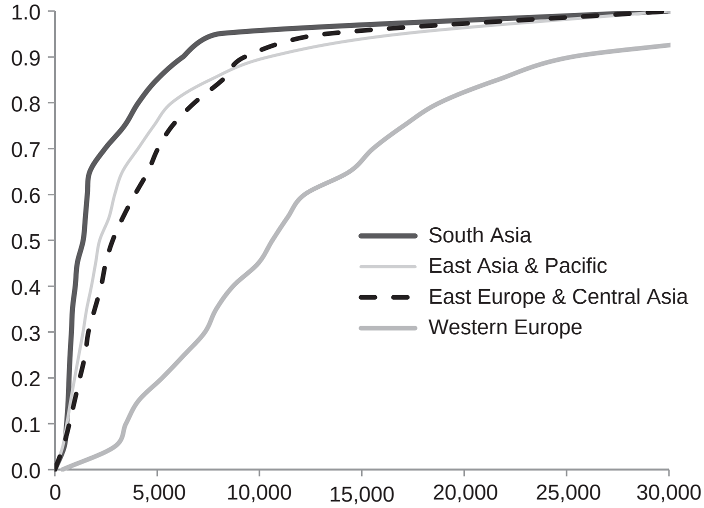


Figure 9.19 Distribution functions. *Note: Cumulative distribution functions of per capita household income. Source: Own estimates based on microdata from Gallup World Poll, 2006.*

between 1 and 2 points in MENA and LAC, the size of the effort is larger in Asia and orders of magnitude greater in sub-Saharan Africa.

International surveys, such as the Gallup Poll, provide an opportunity to alleviate some of the typical comparability problems of household surveys because survey design and questionnaires are identical across countries. However, as discussed earlier, these surveys have still small samples, and measurement errors are presumably large, given that only one income question is included. The correlation between headcount ratios computed with the Gallup Poll and PovcalNet is 0.32, significant at 2%, whereas the rank Spearman correlation is 0.61, significant at 1%.

Figure 9.19 shows the cumulative density function in some regions of the world, based on Gallup data. There is first-order stochastic dominance of the Western Europe distribution over the rest, whereas the South Asian distribution is dominated by the rest.[100] The curves of ECA and EAP cross each other, although they do so at high-income levels.

9.6. POVERTY: TRENDS

In the last decades the developing world has made undeniable progress toward the goal of ending absolute poverty. The evidence suggests that the first goal of the Millennium Development Goals—to halve extreme poverty from 1990 to 2015—was already met in 2010 for the aggregate of developing countries. However, the strong decline in global absolute poverty hides substantial heterogeneities across economies and over time. In this section we document and characterize trends in income poverty in the countries of the developing world since the early 1980s to 2010 and trace those changes to economic growth and distributive changes.[101]

The literature on international poverty trends can be divided into two main strands. The first one makes comparisons based exclusively on household survey microdata. This ambitious road that requires access to surveys for many countries over time has been taken by several initiatives at the regional level and by the World Bank at a global scale, mainly through the work of Martin Ravallion and Shaohua Chen. The second strand combines estimates of the national income distributions with GDP or aggregate consumption data drawn from National Accounts to anchor the mean. Bourguignon and Morrison (2002), Bhalla (2002), Karshenas (2003), and Sala-i-Martin (2006) are examples of this literature. In this section we mainly document the results of the first approach.

9.6.1 Trends in Income Poverty

Data from PovcalNet reveal a widespread fall in absolute income poverty in the developing countries over the last three decades (Figure 9.20). Only for a few nations poverty in 2010 was not lower than in 1981; that set includes some economies in Eastern Europe,

[100] First-order stochastic dominance ensures that the result of lower poverty in Western Europe is robust to the choice of poverty line and valid for a broad class of poverty measures (Atkinson, 1987).

[101] Chapter 11 in this volume also covers the issue of global poverty in the developing world. In our chapter the interest is more focused on the country trends, and then we make more use of unweighted statistics.

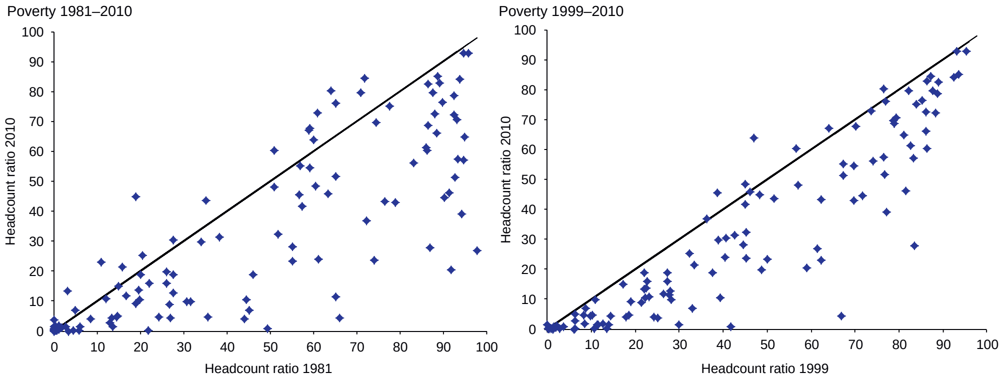


Figure 9.20 Poverty headcount ratio. Developing countries. *Note: Poverty computed over the distribution of consumption/income per capita with the PPP-adjusted \$2-a-day line. Source: Own calculations based on PovcalNet (2013).*

Central Asia, and sub-Saharan Africa and only a few in Latin America and the Caribbean. The poverty decline in the 2000s was even more generalized: in only 8 out of 121 countries did poverty increase between 1999 and 2010 (5 in sub-Saharan Africa).

The proportion of the population in the developing world living with less than \$1.25 per day decreased from 52% in 1981 to 20.8% in 2010, which implies a trend decline of around one point per year (Figure 9.21).[102] This is a remarkable achievement that should not be overlooked. It would be hard to find other episodes in history where the extent of extreme poverty was reduced so massively in such a short period of time.

However, this extraordinary result should be put in perspective. First, even after this global social improvement one of every five persons in the developing world still lives in extremely deprived conditions (less than \$1.25 a day), whereas four out of 10 people have household per capita consumption levels lower than a frugal \$2 per day.

Second, the positive performance of China is key for the global result. Ignoring China, the poverty decline is less impressive: the headcount ratio fell 16 instead of 31 points in three decades (Table 9.14).[103] In fact, when ignoring China, it is not clear that the developing world could meet the MDG for poverty reduction in 2015. Although the population-weighted mean of the poverty rate dropped 31 points between

[102] The cumulative distribution function of 2010 lies always below the functions corresponding to all previous years (first-order stochastic dominance).

[103] When ignoring China the distribution of 2010 still dominates (first-order stochastic dominance) 1981, although the distance between the two cumulative distributions is smaller. The curve for 1999 lies below 1981 and 1990 but only for poverty lines lower than \$3 a day.

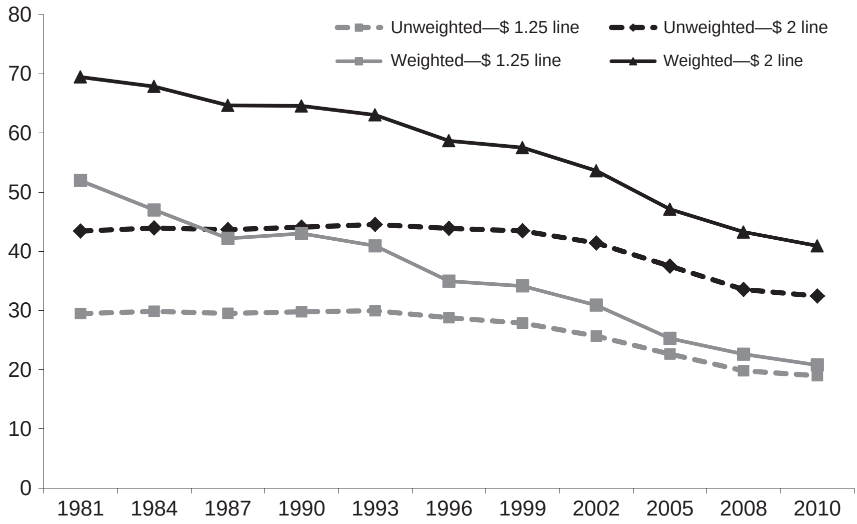


Figure 9.21 Poverty headcount ratio. Weighted and unweighted means, developing countries, 1981–2010. *Note: Poverty computed over the distribution of consumption/income per capita. Source: Own calculations based on PovcalNet (2013).*

1981 and 2010, the unweighted mean declined about 10 points, and the median went down just 8 points. The extreme poverty rate ($1.25) of a typical developing country was reduced from 29.5% in 1981 to 19% in 2010, which represents a fall of around a third of a point per year, a figure less impressive than the corresponding one for the global poverty rate (one point a year). In fact, this decline took place only since the late-1990s. A typical developing country did not experience any improvement for almost two decades: the unweighted poverty rate for the developing world was 29.5% in 1981, 29.8% in 1990, and 28.8% in 1996. From then on poverty declined more consistently, especially between 2002 and 2008, when for a typical developing economy the rate of poverty reduction was almost one point a year.

The fall in poverty appears less startling when using higher poverty lines. Whereas the headcount ratio with the $1.25 line fell 60% from 1981 to 2010, it declined 41% when measured with the $2 line and 20% with the $4 line. In fact, although the MDG goal of halving $1.25-a-day poverty from the value in 1990 was already met in 2010, the assessment is different when using the $2 line: the headcount ratio in 2010 was about two-thirds of the value in 1990.

The performance of the developing world in terms of poverty reduction also looks much less spectacular when considering poverty counts, instead of the typical measures that are invariant to the size of the population. In fact, for the majority of the developing countries (63%) the number of poor people was higher in 2010 than in 1981. Even during the booming 2000s the poverty count increased in 30% of the nations. The number

Table 9.14 Change in Poverty Measures (Points)
Developing countries, 1981–2010

Poverty measure	1981–1990	1990–1999	1999–2010	1981–2010	Poverty measure	1981–1990	1990–1999	1999–2010	1981–2010
Headcount—$1.25 line					**Poverty gap—$1.25 line**				
Unweighted					***Unweighted***				
Mean	0.3	−1.9	−8.9	−10.5	Mean	0.2	−1.6	−4.2	−5.7
Median	4.9	0.0	−12.9	−8.0	Median	0.3	0.3	−4.0	−3.4
Weighted					***Weighted***				
Mean	−9.0	−8.9	−13.4	−31.2	Mean	−6.4	−3.7	−4.8	−14.9
Without China	−2.8	−3.1	−10.0	−15.9	Without China	−1.6	−1.5	−3.7	−6.8
Headcount—$2 line					**Poverty gap—$2 line**				
Unweighted					***Unweighted***				
Mean	0.7	−0.6	−11.0	−11.0	Mean	0.3	−1.5	−6.5	−7.7
Median	−0.3	−3.4	−17.1	−20.7	Median	2.3	0.8	−9.2	−6.1
Weighted					***Weighted***				
Mean	−4.9	−7.1	−16.6	−28.5	Mean	−6.6	−5.4	−8.8	−20.8
Without China	−1.2	−1.1	−10.9	−13.1	Without China	−1.7	−1.8	−6.4	−9.9
Headcount—$4 line					**Poverty gap—$4 line**				
Unweighted					***Unweighted***				
Mean	0.9	2.2	−11.4	−8.3	Mean	0.4	−0.4	−9.0	−9.0
Median	2.1	−2.1	−16.3	−16.3	Median	−0.4	−1.4	−12.3	−14.1
Weighted					***Weighted***				
Mean	−1.7	−1.3	−14.0	−17.0	Mean	−8.4	−4.6	−12.1	−25.1
Without China	−1.2	1.7	−8.3	−7.8	Without China	−1.4	−0.5	−7.8	−9.7

Note: Poverty computed over the distribution of consumption/income per capita.
Source: Own calculations based on PovcalNet (2013).

of poor people was in the late 2000s only slightly lower than in the early 1980s (2585 million in 1981 and 2394 million in 2010 with the $2 line); ignoring China, the poverty count actually increased in 422 million (from 1613 to 2035). Some authors have suggested that the difference in the assessment of world poverty when alternatively using the headcount ratio or the number of poor people may be behind some of the controversies in the public debate about globalization and the social performance of the world in the last decades.[104]

Because the calculation of global poverty with microdata is very cumbersome and requires having a large number of comparable household surveys, some authors estimate the changes in poverty with National Accounts data, anchoring the mean of the income distribution to output or domestic consumption taken from National Accounts, using secondary distributive data and making assumptions about the functional form of the income distribution, typically the lognormal parameterization (Ahluwalia et al., 1979; Bhalla, 2002; Chotikapanich et al., 1997; Pinkovskiy and Sala-i-Martin, 2009; Sala-i-Martin, 2006).[105] This methodology allows ambitious calculations in terms of coverage, but it faces some obvious caveats as changes in National Accounts aggregates may be misleading proxies for changes in household per capita income, and the secondary distributive data in which the estimates are usually based are mined with several comparability problems.[106]

According to the estimates by Pinkovskiy and Sala-i-Martin (2009) (PSM) using a sample of 191 countries, and distributive data from the WIID, world poverty rates ($2 line) went down from 45.2% in 1970 to 37.8% in 1981, 24.9% in 1990, 16.8% in 1999, and 13% in 2006. That pattern implies a fall substantially faster than what is estimated with household surveys data alone in the 1980s and 1990s, but slower in the 2000s.[107] For instance, although for 2008 the magnitude of the poverty rates in PSM for the $3 line is similar than in PovcalNet with the $1.25 line, the declines have been different. In PSM the headcount ratio fell at annual 2.6% and 3.5% in the 1980s and 1990s, respectively, whereas the rates were 2.1% and 2.5% with PovcalNet data. Instead, in the 2000s poverty fell at annual 3.1% in PSM and at annual 4.4% in PovcalNet.

[104] See Reddy and Pogge (2010), Chen and Ravallion (2012), and Cockburn et al. (2012).

[105] It should be noticed that the World Bank poverty estimates also use National Accounts (NAS) data, although to a limited extent. For instance, for countries with only one household survey, poverty is estimated by applying the NAS consumption estimates to the available distribution data, assuming the Lorenz curve remains fixed.

[106] An intermediate alternative was proposed by Karshenas (2003), who calibrated survey means using National Accounts statistics. Calibrated survey means are read off the fitted curve for the mean of household survey consumption conditional on the NAS mean.

[107] Dhongde and Minoui (2010) explore several factors behind the different results in Chen and Ravallion (2010) and Pinkovskiy and Sala-i-Martin (2009).

9.6.2 Decomposing Poverty Changes

Researchers frequently use decompositions of changes in poverty into growth and redistribution effects to characterize poverty trends.[108] The growth effect refers to the poverty change between two years that would have occurred if the mean income had changed as it did, but the shape of the distribution had stayed fixed. On the other hand, the redistribution effect records the poverty change between two years that would have occurred if the shape of the distribution had changed in the way it did, but the mean had remained fixed. Of course, this is just a mechanical exercise, as indicators of economic growth and changes in inequality and poverty are just three different ways of aggregate information on income dynamics, and therefore they are all jointly determined by the general equilibrium of the economy. In that sense the decompositions are helpful to illustrate the way incomes have changed and affected poverty, but they are silent on the fundamental factors underlying poverty changes and on the policies recommended to reduce deprivations more effectively.

The change in the growth-inequality-poverty triangle was very different in the 2000s compared to the previous decade. Table 9.15 illustrates this difference showing the unweighted mean of the growth and redistribution effects of changes in poverty for a sample of 76 developing countries with information on deciles shares in PovcalNet.[109] The mild fall in the poverty headcount ratio ($2 line) in the 1990s can be decomposed into a poverty-decreasing growth effect that outweighed a poverty-increasing redistribution effect. On average (unweighted) mean consumption grew at an annual 0.2%, and the Gini increased about 0.3 points per year, implying a very modest decline in poverty. In contrast, in the 2000s both effects contributed to a falling poverty. A robust increase of mean consumption (more than 3% a year) and a slow fall in inequality (around 0.1 Gini

Table 9.15 Decomposition of poverty changes
Developing countries, 1990–2010

		Effects	
	Actual change	**Growth**	**Redistribution**
1990–1999	−0.3	−1.5	1.2
1999–2010	−10.1	−9.0	−1.1
1990–2010	−10.5	−10.4	−0.1

Note: The columns show unweighted means across a sample of 76 developing countries of the change in poverty headcount ratio ($2 line), and the growth and redistribution effects from the poverty change decomposition.
Source: Own calculations based on data from PovcalNet.

[108] See Datt and Ravallion (1992), Kakwani and Subbarao (1990), Kakwani (2000), and Mahmoudi (2001) for different proposals.

[109] Because the decompositions could be carried out changing the base year, the table shows the averages over the two exercises. To obtain the results we use lineal approximations to the Pen curve based on information on mean consumption per capita by decile from PovcalNet.

points a year) combined to yield a substantial reduction in indicators of material deprivation. The growth effect was particularly large, accounting for 90% of the fall in the headcount ratio ($2 line). This result does not mean that distributive changes are not important, but instead that they have not been the main drivers of poverty reduction in the past.

9.6.3 The Cost of Closing the Poverty gap

Although still a challenging problem, eliminating absolute extreme income poverty is an increasingly affordable target. Based on PovcalNet data, we computed the population-weighted poverty gap index in the developing world as a share of global GDP. This fraction, which indicates the fiscal effort required to end poverty in a scenario of perfect targeting, absent behavioral responses, has been substantially falling over time as poverty decreased and global GDP went up. The resources needed to close the poverty gap with the $1.25 line as a share of global GDP declined from 1.3% in 1981 to 0.2% in 2010 (the corresponding values for the $2 line are 3.6% and 1%).[110] As a share of the GDP in the developing world the cost of closing the poverty gap fell from 1.9% in 1981 to 0.4% in 2010 (5.4% and 1.8% for the $2 line).

Kanbur and Sumner (2011) highlight the fact that although in 1990 over 90% of the world's extremely poor people ($1.25) lived in countries classified as low-income countries (LICs), by the late 2000s this share dropped to less than 30%. From the fact that most of the world's poor live in middle-income countries with the domestic financial capacity to end at least extreme poverty, Sumner (2012) concludes that poverty reduction is increasingly becoming a domestic issue of national distribution and domestic political economy, rather than primarily an aid and international issue. Table 9.16 suggests than on average (unweighted) across developing countries the redistributive national effort to end extreme poverty under perfect targeting fell from 8.2% of GDP in 1981 to 4% in 2010. The median value is much lower and has fallen from 1.9% in 1981 to 0.5% in 2010. The third column shows the share of countries where the cost of eliminating extreme income poverty is less than 1 GDP point. That share substantially increased in the 2000s, from 41.3% in 1999 to 55.4% in 2010. Similarly, the share of countries for which closing the poverty gap is fiscally very burdensome (in the table more than three GDP points) fell from more than 50% in 1990 to about 30% in 2010. Sumner (2012) reports similar trends, although a lower proportion of countries with poverty gap/GDP less than 1%. Ravallion (2010) also notes that most middle-income countries would require very small additional taxation to end poverty.

[110] This computation assumes zero poverty in the high-income countries, when deprivation is measured with the international lines.

Table 9.16 Poverty gap as share of GDP
Mean, median, share of countries with gap/GDP less than 1% and greater than 3%
Developing countries, 1981–2010

	Poverty line $1.25				Poverty line $2			
	Mean	Median	% Less 1 GDP point	% More 3 GDP points	Mean	Median	% Less 1 GDP point	% More 3 GDP points
1981	8.2	1.9	42.4	45.7	19.9	5.4	22.9	60.4
1984	8.4	2.2	41.3	45.7	20.5	6.6	19.8	60.4
1987	8.2	2.3	39.1	46.7	19.9	7.1	20.8	60.4
1990	8.1	3.3	39.1	51.1	19.6	7.0	24.0	59.4
1993	9.2	2.4	40.2	47.8	22.0	7.5	25.0	59.4
1996	9.2	2.0	40.2	42.4	21.6	7.3	24.0	60.4
1999	6.9	2.0	41.3	39.1	17.9	5.9	25.0	60.4
2002	5.9	1.4	41.3	40.2	15.7	4.8	27.1	58.3
2005	5.2	1.0	48.9	35.9	14.0	3.7	32.3	55.2
2008	4.3	0.7	50.0	33.7	11.6	3.0	40.6	50.0
2010	4.0	0.5	55.4	31.5	11.1	3.0	42.7	50.0

Note: Poverty computed over the distribution of consumption/income per capita. Unweighted statistics.
Source: Own calculations based on PovcalNet (2013).

9.6.4 Regional Trends

The patterns in income poverty over time have been heterogeneous across the geographic regions of the developing world (Table 9.17). At least three facts are worth stressing: (i) the remarkable decline in poverty in Asia over the last three decades, (ii) the lack of significant progress in the rest of the regions in the 1980s and 1990s, and (iii) the generalized fall in poverty in the 2000s.

Arguably, the main fact about poverty dynamics in the last three decades has been the noteworthy decline in Asia. The share of people living with less than $2 a day went down from 92.4% to 29.7% between 1981 and 2010 in East Asia and Pacific, and from 87.2% to 66.7% in South Asia. The performance of EAP is enhanced by the presence of China, but also the unweighted mean dramatically fell in this region (from 70.4% in 1981 to 38.4% in 2010). The fall in the unweighted mean was similar in South Asia (from 80.3% to 43.5%). Almost all countries in Asia experienced drops in income poverty over the period 1981–2010. The reductions were on average larger in the 2000s than in the previous decades. For instance, in South Asia the unweighted mean fell 5.6 points in the 1980s, 10.9 in the 1990s, and 20.4 in the 2000s.

In the rest of the developing world the performance was weak and even negative in the 1980s and 1990s. Income poverty rose in Latin America in the 1980s, in Eastern Europe and Central Asia in the 1990s, and in sub-Saharan Africa in both decades. In contrast, all regions experienced falling poverty in the 2000s. The reductions were large and in general outweighed the weak performance of the previous decades. For instance, on

Table 9.17 Poverty headcount ratio, developing world 1981–2010
Weighted and unweighted means

	1981	1990	1999	2010
Unweighted mean; $1.25 line				
East Asia and Pacific	53.4	44.3	33.3	17.9
Eastern Europe and Central Asia	4.0	4.5	7.5	1.4
Latin America and the Caribbean	13.2	14.4	13.7	8.6
Middle East and North Africa	10.3	7.2	6.8	3.5
South Asia	59.7	51.3	37.8	17.6
Sub-Saharan Africa	49.3	54.0	52.2	41.5
Unweighted mean; $2 line				
East Asia and Pacific	70.4	65.8	56.2	38.4
Eastern Europe and Central Asia	10.1	10.2	17.0	4.8
Latin America and the Caribbean	24.3	25.9	23.8	15.6
Middle East and North Africa	28.6	24.0	23.5	13.5
South Asia	80.3	74.7	63.9	43.5
Sub-Saharan Africa	67.1	71.5	71.0	62.2
Population-weighted mean; $1.25 line				
East Asia and Pacific	77.2	56.2	35.6	12.5
Eastern Europe and Central Asia	1.9	1.9	3.8	0.7
Latin America and the Caribbean	11.9	12.3	11.9	5.5
Middle East and North Africa	9.6	5.7	5.0	2.4
South Asia	61.1	53.8	45.1	31.0
Sub-Saharan Africa	51.4	56.5	57.9	48.5
Population-weighted mean; $2 line				
East Asia and Pacific	92.4	81.0	61.7	29.7
Eastern Europe and Central Asia	8.3	6.9	12.1	2.3
Latin America and the Caribbean	23.7	22.5	22.0	10.4
Middle East and North Africa	30.0	23.4	21.9	12.0
South Asia	87.2	83.6	77.8	66.7
Sub-Saharan Africa	72.2	75.9	77.5	69.9

Source: Own calculations based on PovcalNet (2013).

average (unweighted) income poverty ($2 line) went down 72% in ECA, 43% in MENA, 34% in LAC, and 12% in SSA. The proportion of countries where the headcount ratio fell more than 5% in the 2000s is above 90% in all these regions, with the exception of SSA, where the corresponding proportion is 66%.

The contrast between Asia and the rest of the developing world in terms of poverty reduction is even more dramatic when the calculations are carried out anchoring the mean of the income distribution to GDP from National Accounts. Figure 9.22 shows regional estimates taken from Pinkovskiy and Sala-i-Martin (2009), where the sharp

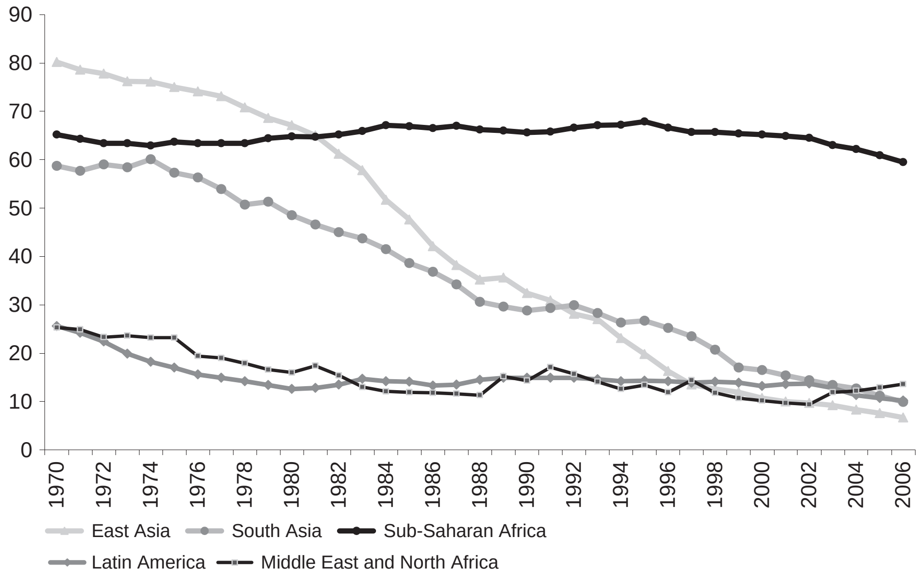


Figure 9.22 Poverty headcount ratio. Developing countries, 1970–2006. *Note: Poverty line $2 a day. Source:* *Pinkovskiy and Sala-i-Martin (2009).*

declines in poverty in East and South Asia stand out. According to these estimates there would be poverty convergence across the regions of the developing world, with the exception of sub-Saharan Africa.

Regions have been different in terms of the growth-inequality-poverty triangle (Table 9.18). The growth effect was strong in Asia dwarfing a much smaller (and sometimes poverty-increasing) redistribution effect. The increase in poverty in ECA in the 1990s is associated with both negative growth and inequality rise, whereas the fall in poverty in the following decade is mostly accounted for by neutral positive growth. In Latin America growth contributed in both decades, but only in the 2000s did the redistribution effect become poverty-reducing. Finally, in Africa (MENA and SSA) the growth effect in the 2000s was the major contributing factor to the fall in poverty.

In the rest of this section we take a closer look at the changes in poverty reported by the literature in each geographic region of the developing world.

9.6.4.1 *East Asia and Pacific*

As mentioned earlier, East Asia and Pacific has achieved an impressive record in terms of poverty reduction. The fall in the indicators of material deprivation has been strong, sustained over the two last decades and widespread across countries.[111] The poverty headcount ratio with the $2 line fell from 92.4% in 1981 to 81% in 1990, 61.7% in 1999, and 29.7% in 2010.

[111] See Ahuja et al. (1997) and Nissanke and Thorbecke (2010) as examples of a vast literature.

Table 9.18 Decomposition of poverty changes
Developing countries

	Actual change	Effects: Growth	Effects: Redistribution		Actual change	Effects: Growth	Effects: Redistribution
EAP				**MENA**			
1990–1999	−10.5	−10.7	0.2	1990–1999	−1.5	−0.6	−1.0
1999–2010	−19.4	−19.2	−0.2	1999–2010	−10.1	−8.0	−2.1
1990–2010	−30.0	−30.3	0.3	1990–2010	−11.6	−8.4	−3.3
ECA				**SA**			
1990–1999	8.6	6.1	2.5	1990–1999	−9.0	−10.9	2.0
1999–2010	−11.5	−10.5	−1.0	1999–2010	−13.2	−12.0	−1.1
1990–2010	−2.9	−3.9	1.0	1990–2010	−22.1	−23.3	1.2
LAC				**SSA**			
1990–1999	−3.2	−5.3	2.1	1990–1999	−1.8	−1.4	−0.4
1999–2010	−7.9	−5.3	−2.6	1999–2010	−5.9	−6.0	0.0
1990–2010	−11.1	−10.2	−0.9	1990–2010	−7.7	−7.8	0.0

Note: The columns show the unweighted means of the change in the poverty headcount ratio and the growth and redistribution effects from the poverty change decomposition. They are based on a subsample of 76 countries in PovcalNet for which all data needed for the decomposition were available.
Source: Own calculations based on data from PovcalNet.

China's progress against absolute poverty was a key factor behind this dynamic (Minoiu and Reddy, 2008; Ravallion and Chen, 2007; World Bank, 2009). Rural areas accounted for the bulk of the gains to the poor, although migration to urban areas helped; agriculture played a far more important role than the secondary or tertiary sources of GDP, mainly from the efficiency gains after the decollectivization process. Ravallion and Chen (2007) claim that "the halving of the national poverty rate in the first few years of the 1980s was largely attributable to picking these low-lying fruits of agrarian reform." Provinces starting with relatively high inequality saw slower progress against poverty, due both to lower growth and a lower growth elasticity of poverty reduction. In 1990 the incidence of poverty in China was roughly 25 points higher than in the rest of the developing world, but by the end of the 2000s, it had fallen more than 10 points below the average.[112]

Between 1990 and 2010 the headcount ratio fell from 60% to 12% in China ($1.25 line), and from 54% to 18% in Indonesia, the two most populated countries in the region.

[112] During the early and mid-1990s poverty declined substantially, but then in the late 1990s to the early 2000s the downward trend stalled. Li et al. (2013) argue that further reductions in poverty became more challenging due to several factors, for example, the fact that a high proportion of the remaining poverty was geographically dispersed and transient and also because poverty had become less responsive to macroeconomic growth (World Bank, 2009). Policies adopted after 2002, such as the minimum living guarantee program, the new rural cooperative medical system, and the new rural pension system, have addressed some of these factors.

The rest of the East Asian economies experienced similar patterns. The headcount ratio dropped from 58% to 15% in Cambodia, from 12% to 0.4% in Thailand, and from 73% to 14% in Vietnam. The reductions were less spectacular, but anyway significant in the Pacific countries (Micronesia, Philippines, Papua New Guinea, Timor-Leste). According to the Asian Development Bank (2012a, 2012b), and using the national lines, poverty declined in all EAP economies except Mongolia, Micronesia, Samoa, Timor-Leste, Tonga, and Tuvalu. With respect to the poverty reduction target of the Millennium Development Goals, of the 10 economies for which data are available, six have achieved the poverty target, and Cambodia is very close to reaching it by 2015. Lao PDR, Philippines, and Timor-Leste are progressing more slowly toward that goal.

9.6.4.2 Eastern Europe and Central Asia

The evidence clearly suggests that poverty in Eastern Europe and Central Asia increased during the 1990s and decreased during the 2000s. The collapse of output in many of these countries following the dismantling of the Soviet Union, along with hyperinflation that wiped out savings, resulted in a dramatic drop in living standards for the majority of people and the emergence of poverty as a major issue.[113] However, according to Simai (2006), poverty in the region was not a new phenomenon; most of the countries began their transformation with extensive hidden unemployment and at least one-tenth of its population below the subsistence level. The transition also involved setbacks in non-monetary dimensions of well-being. Past achievements in social welfare came under pressure, the most striking being the reduction in life expectancy: In 1995, life expectancy for males in the Russian Federation was just 58 years, 10 years less than that of men in China.

Table 9.17 reminds the reader that poverty rates in these countries have always been much lower than in the other developing regions, irrespective of the line considered and of the weighted or unweighted averages, and despite the fact that low-income CIS (Georgia, Uzbekistan, Moldova, Armenia, Kygryz Republic, Tajikistan) display figures well above the regional mean. When growth resumed after the Russian crisis in 1998, poverty started to fall.[114] It was not until 2004 that the region as a whole returned to the level of GDP recorded in 1990 (World Bank, 2005). Much of the poverty reduction initially occurred in the populous middle-income countries (Kazakhstan, Russian Federation, Ukraine), but eventually it spread almost everywhere. Scholars have identified four main (nonindependent) explanations for the poverty dynamics: (i) the positive growth in output and wages; (ii) the decline in inequality, with incomes of poorer households increasing faster than average in some countries (CIS) but not in others (Poland,

[113] Poverty in Eastern Europe and the Soviet republics pre-1990 is discussed at some length in Atkinson and Micklewright (1992).

[114] For an analysis of the growth elasticities of poverty reduction, see World Bank (2005), Chapter 1.

Romania); (iii) the increasing role of public transfers, with benefits improving in coverage and adequacy; and (iv) private remittances, which in many cases far exceeded public funding.

9.6.4.3 Latin America and the Caribbean

Poverty estimates in Latin America, available since the 1970s, were always mostly based on income, not consumption.[115] Despite the fact that statistics were initially weak, there is agreement among researchers that during the 1970s economic growth pushed poverty down in the region (Altimir, 1979, 1996). In contrast, the "lost decade" of the 1980s was characterized by recurrent crises and output stagnation, which brought about a weak social performance. Latin American economic growth resumed in the 1990s, but in the context of growing inequalities, a combination that resulted in a modest decline in poverty (Londoño and Székely, 2000). The exceptional economic conditions in the 2000s and the consolidation of more ambitious social policies implied a sharp drop in poverty indicators in that decade. Gasparini et al. (2013) estimate that the income poverty headcount ratio ($2.5 line) fell from 36% in the early 1970s to 27.3% in 1980, slowly increased to reach 27.8% in 1992, mildly decreased to 24.9% in 2003, and was reduced dramatically to 16.3% by 2009.[116] The number of poor people in the region is estimated to have fallen from 119 million in 1992 to 89 million in 2009 (Gasparini et al., 2013). Other indicators of income poverty and of various types of material deprivation are consistent with these results. Data from SEDLAC confirms that the reduction in poverty continued in the period 2009–2013, despite a deceleration in the GDP growth rates. The improvements in social indicators in the 2000s have been linked to at least two factors: on the one hand, most of the region's economies experienced robust growth together with upswings in employment and labor income; on the other, all countries boosted social spending and put ambitious social protection systems into place or expanded the scope of their existing systems (Cruces and Gasparini, 2012; López Calva and Lustig, 2010).

The performance in terms of poverty reduction has been heterogeneous across the regions in Latin America. Although Central America experienced a slow decline in income deprivation over the last two decades, in the rest of the regions the rapid poverty fall of the 2000s contrasts with the sluggish or even negative performance of the 1990s.[117]

[115] In the Caribbean poverty is estimated based on consumption expenditures, but surveys are still scarce, so poverty changes are difficult to monitor.

[116] Estimates are based on SEDLAC statistics, Londoño and Székely (2000), Wodon et al. (2001), and official poverty estimates from all countries in Latin America.

[117] SEDLAC data indicate that using the international poverty line of $2.5, the average (unweighted) poverty rate in Southern South America increased from 17.7% in 1992 to 18.5% in 2003, and then dropped to 9.1% by 2010. In the same period the average poverty rate in the Andean region first rose from 30.2% to 33%, and then declined to 17.5%. In contrast, poverty in Central America fell more uniformly over the two decades: 33.3% in 1992, 28.6% in 2003, and 23.1% in 2010.

More than half of the population in Latin America lives in Brazil or Mexico. After a decline in the early 1990s, income poverty in Brazil remained stable for about a decade; the poverty rate with the $2.5 line was 27.8% in 1995 and 27.1% in 2003. After that plateau, the country experienced a fast and sustained reduction in income poverty, reaching 13.9% in 2010. The Mexican economy was hit by a serious crisis in the mid-1990s (the so-called Tequila crisis) that raised income poverty from 17.8% in 1992 to 43.4% in 1995 (SEDLAC data, $2.5 line). From that peak, income poverty initiated a consistent decline to reach a value of 12% in 2006, with no gains in the second half of the 2000s.

9.6.4.4 Middle East and North Africa

As discussed in Section 9.4, it is useful to divide the last four decades in MENA into three periods. The first one, spanning until the mid-1980s, was characterized by strong economic growth: the average per capita income growth for the region between 1975 and 1984 was over 4.5%. Adams and Page (2003) argue that given MENA´s relatively equal income distribution, this economic growth had a powerful impact on reducing poverty in the region. However, assessing that progress is difficult becaused pre-1985 household surveys are only available for Tunisia and Egypt. Iqbal (2006) reports that in Tunisia poverty fell from 51% in 1965 to 16% in 1985, while Egypt's achievement was also impressive, with poverty declining from 82% to 53% between 1975 and 1985. Page (2007) estimates a fall in the incidence of poverty ($2 line) in the region from 57% in the late 1970s to 22% in the early 1990s.

The second period, covering the late 1980s and most of the 1990s, was characterized by a sluggish economic performance, in part due to low prices for hydrocarbons, declining remittances and aid flows, as well as a low payoff to the reforms that were implemented. On average for the developing MENA countries, real per capita incomes went up by less than 1.5% per year. The proportion of those living with less than $2 per day stayed roughly unchanged at around 22% for a decade (PovcalNet). Iqbal (2006) reports that "by 2001, approximately 52 million people were poor, an increase in absolute numbers of approximately 11.5 million people, compared with the situation in 1987." The 1990s were a lost decade for economic growth and poverty reduction in the developing economies of the MENA region (Page, 2007).

The downward pattern in poverty appeared to have resumed in the 2000s. The incidence of poverty ($2 line) fell from 22% in 1999 to 12% in 2010. According to PovcalNet data the number of people living with less than $2 a day increased by 7 million in the 1990s and then was reduced by 20 million in the 2000s (from 60 to 40 million). In the 1990s poverty went down in a third of the MENA countries, but in the 2000s it went down in all economies, with the possible exception of Yemen.

9.6.4.5 South Asia

South Asia achieved impressive economic growth in the past 15 years. Since 1996 until the recent global crisis, GDP growth in the region exceeded 5% per year. As a consequence, poverty rates were considerably reduced. In Bangladesh, India, and Nepal, absolute poverty (headcount ratio, $1.25 line) fell by 18, 15, and 43 points, respectively, between 1996 and 2010. In Pakistan, it fell by 22 points between 2002 and 2010. Devarajan and Nabi (2006) optimistically expect the region to have single-digit poverty rates in 2015 if growth rates accelerated to 10% per year. This has indeed happened already in Sri Lanka (where poverty dropped from 17 in 1996 to 4 in 2010) as well as in Maldives (where, according to the 2006 census, about 1 of the population was living on less than $1 a day).

Growth has been instrumental in reducing poverty rates, but as Ghani (2010) points out from a gloomier perspective, poverty rates were not falling fast enough to reduce the total number of poor people. Those living on less than $1.25 a day increased from 549 million in 1981 to 595 million in 2005. In India, where around three-quarters of these poor live, the number increased from 420 million in 1981 to 455 million in 2005 (Ravallion et al., 2009). The situation seems to have slightly improved according to the most recent observations.

9.6.4.6 Sub-Saharan Africa

The economic and social performance of sub-Saharan African countries has been frustrating. Five decades after most nations gained independence, poverty is still at very high levels in SSA, in fact the highest in the world. Fortunately, it seems that finally, after years of impoverished economic performance, the last decade shows some signs of improvement. Based on figures from PovcalNet, the incidence of poverty increased between 1981 and 1999 from 51% to 58% for the $1.25 per day line and from 72% to 77% for the $2 line (the unweighted statistics are not very different).[118] Because of the increase in population the number of people living with less than $1.25 a day almost doubled in sub-Saharan Africa during those years, from 205 million to 377 million. These results are even more disappointing when compared to the rest of the developing world. The share of the world's poor living in sub-Saharan Africa rose from 11% in 1981 to 22% in 1999. Artadi and Sala-i-Martin (2003) find that the SSA poverty rate was stable in the 1970s, around a value of 49%, and soared in the 1980s and early 1990s, reaching 60% in 1995, following the deterioration of aggregate per capita income in the region. They report a nearly 50% reduction in the worldwide poverty rate between 1980 and 2000, which contrasts with a 27% increase in SSA.

Following the dismal performance of the 1980s and early 1990s, SSA witnessed some economic and social progress. Poverty declined considerably from their heights of the

[118] According to the $4 line, more than 90% of the SSA population is actually considered poor.

mid-1990s; in fact the reduction in extreme poverty from that date to the late 2000s was similar than in the rest of the developing world, excluding China. Chen and Ravallion (2012) stressed the fact that for the first time since 1981 the share of people in sub-Saharan Africa living below \$1.25 a day was less than 50%. Changes in poverty have been closely linked to economic growth. According to Fosu (2010), poverty has declined in SSA by about 0.5 points per year since the mid-1990s, quite comparable with South Asia's record. Sala-i-Martin and Pinkovskiy (2010) estimate that African poverty has been falling steadily since 1995. According to these authors the African poverty rate in 2006 was 31.8%, 30% lower than in 1995, and 28% lower than in 1990. The decline in poverty in the 2000s was quite extended: As reported earlier, two-thirds of the SSA countries in PovcalNet experienced reductions larger than 5%. Sala-i-Martin and Pinkovskiy (2010) find that "poverty fell for both landlocked as well as coastal countries; for mineral-rich as well as mineral-poor countries; for countries with favorable or with unfavorable agriculture; for countries regardless of colonial origin; and for countries with below or above median slave exports per capita during the African slave trade. Hence, the substantial decline in poverty is not driven by any particular country or set of countries." However, Fosu (2009) highlights the considerable heterogeneity across country experiences. For example, although high economic growth in Botswana has been transformed to only a minimal decline in poverty, Ghana has succeeded in translating its relatively modest growth to considerable poverty reduction. Fosu (2009) attributes this contrast to the difference in the levels of income inequality between the two countries.

Despite the encouraging signs of progress in the fight against poverty in sub-Saharan Africa, Chen and Ravallion (2012) alert that the lags in survey data availability and problems of comparability and coverage raise concerns about how robust this positive trend will prove to be.

9.6.5 Poverty Convergence

Given the heterogeneity among economies in terms of social improvement, a natural question is whether countries starting out with a high incidence of material deprivation tend to be the ones with higher rates of poverty reduction; i.e., whether there is *poverty convergence* (Ravallion, 2012).[119] Figure 9.23 suggests some signs of poverty convergence in absolute terms but not in proportionate terms.[120] The 1981 poverty level (\$2 line) is negatively correlated with the annualized change over the period 1981–2010, but not with the proportionate change (annualized log difference). In the first panel the regression line has a slope of −0.009 with a *t*-ratio of −4.51, based on a robust (White) standard

[119] The use of the term *poverty convergence* in this context is not entirely transparent. Another (more trivial) alternative would be assessing whether poverty measures converge toward zero, simply checking that poverty changes are negative.

[120] The analysis is restricted to countries with initial poverty above 5%.

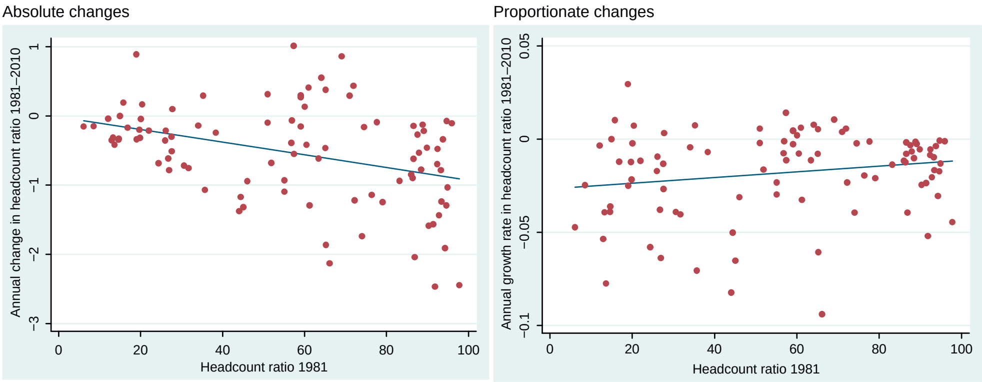


Figure 9.23 Poverty convergence among developing countries 1981–2010. *Note: Poverty computed over the distribution of consumption/income per capita with the PPP-adjusted $2-a-day per person line. Source: Own calculations based on PovcalNet (2013).*

error. This result means that on average the absolute fall in poverty has been larger in countries with higher poverty incidence, but because the difference across countries is small, there is absence of poverty convergence in proportionate terms.

The results for other poverty indices, lines, and time periods are similar (Table 9.19). The coefficients for the absolute change in poverty are negative but small, whereas the coefficients for the proportionate change are in general positive, although in most cases non-significant.[121]

Ravallion (2012) argues that the lack of poverty convergence (in proportionate terms), despite evidence for mean convergence and for the poverty-reduction impact of growth, suggests that the dynamic processes for growth and poverty reduction depend directly on the initial level of poverty. He shows evidence on two adverse effects of being a country with high poverty levels: first, these countries tend to grow at a slower pace, controlling for the initial mean, and second, a high poverty rate implies lower "productivity" of economic growth in terms of poverty reduction (lower poverty-growth elasticity).

9.6.6 Poverty and Growth

The dynamics of poverty are closely related to income growth. The economic literature has gathered abundant evidence supporting the fact that absolute poverty tends to fall

[121] The first panel in Figure 9.23 suggests an increase in the dispersion of absolute changes at higher values of the initial poverty level. A quantile regression analysis confirms that the slope of the regression line is negative and significant for the lower quantiles (the high-performers in terms of poverty reduction), and statistically non-significant for the higher quantiles. The systematic difference in the slopes across quantiles suggests that poverty convergence depends on factors beyond the initial poverty level.

Table 9.19 Poverty convergence among developing countries
Coefficients of initial poverty on a poverty change regression

		1981–2010		1990–2010		1999–2010	
Index	**Poverty line**	**Absolute changes**	**Proportionate changes**	**Absolute changes**	**Proportionate changes**	**Absolute changes**	**Proportionate changes**
Headcount	$1.25	−0.017	0.001	−0.018	0.010	−0.016	0.023
		(0.002)***	(0.004)	(0.003)***	(0.004)*	(0.004)***	(0.006)***
Headcount	$2	−0.009	0.006	−0.008	0.016	−0.008	0.028
		(0.002)***	(0.004)	(0.002)***	(0.004)**	(0.004)**	(0.007)***
Poverty gap	$1.25	−0.021	−0.007	−0.026	−0.001	−0.024	0.014
		(0.002)***	(0.004)	(0.003)***	(0.005)	(0.005)***	(0.007)**
Poverty gap	$2	−0.016	0.001	−0.018	0.010	−0.016	0.024
		(0.002)***	(0.004)	(0.003)***	(0.005)*	(0.004)***	(0.008)***
Squared	$1.25	−0.024	−0.013	−0.032	−0.008	−0.030	0.006
poverty gap		(0.002)***	(0.004)**	(0.004)***	(0.006)	(0.007)***	(0.006)
Squared	$2	−0.020	−0.005	−0.024	0.002	−0.022	0.014
poverty gap		(0.002)***	(0.004)	(0.003)***	(0.004)	(0.005)***	(0.005)***

Note: Standard errors in parenthesis.
*Significant at 10%, **significant at 5%, ***significant at 1%
Source: Own calculations based on PovcalNet (2013).

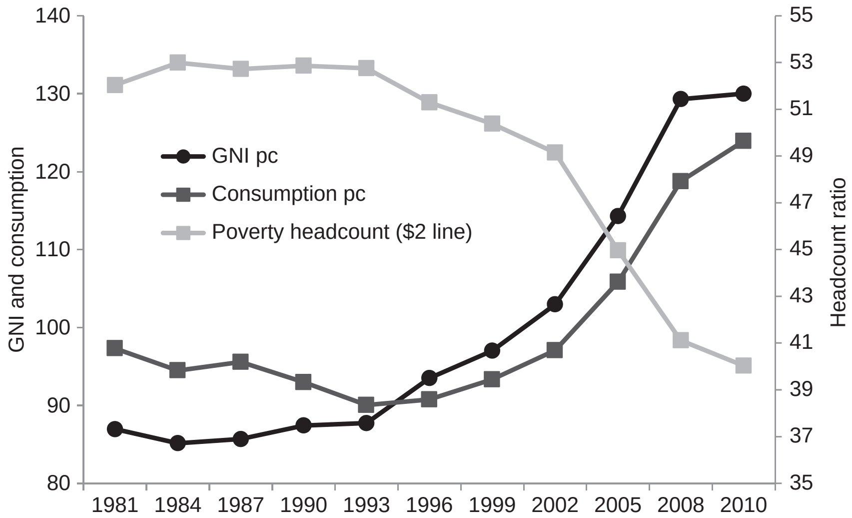


Figure 9.24 Per capita GNI, per capita consumption, and poverty headcount ratio. Unweighted mean, developing countries, 1981–2010. *Note: Per capita gross national income in constant 2005 PPP dollars, per capita consumption from household surveys in constant 2005 dollars. Unweighted averages across developing countries, series displayed with mean = 100. Headcount poverty ratio $2-a-day line, unweighted averages across developing countries. Source: Own calculations based on PovcalNet (2013).*

with economic growth.[122] Moreover, the longer the growth spells under consideration, the larger the share of the variance in poverty that is accounted for by the growth component (Ferreira, 2010). Figure 9.24 illustrates this relationship by showing poverty along with two alternatives measures of mean income: per capita gross national income (GNI) constructed from National Accounts, and per capita consumption obtained from household surveys. The figure shows the unweighted mean of these variables among developing countries in the period 1981–2010. On average, the economic performance of the developing countries was weak in the 1980s and early 1990s, hindering the perspectives of poverty reduction. Since the mid-1990s economic growth resumed in most countries in the developing world, a process that accelerated in the 2000s, leading to a sustained decrease in all measures of income poverty. At the end of the decade that pattern was slowed down, but not halted, by the poor economic performance of several developing countries associated to the international crisis.

[122] See Chen and Ravallion (1997), World Bank (2000), Ravallion (2001, 2007, 2012), Dollar and Kraay (2002), Fields (2002), Bourguignon (2003), Besley and Burgess (2003), Kraay (2006), and Fosu (2011) as examples of a large literature.

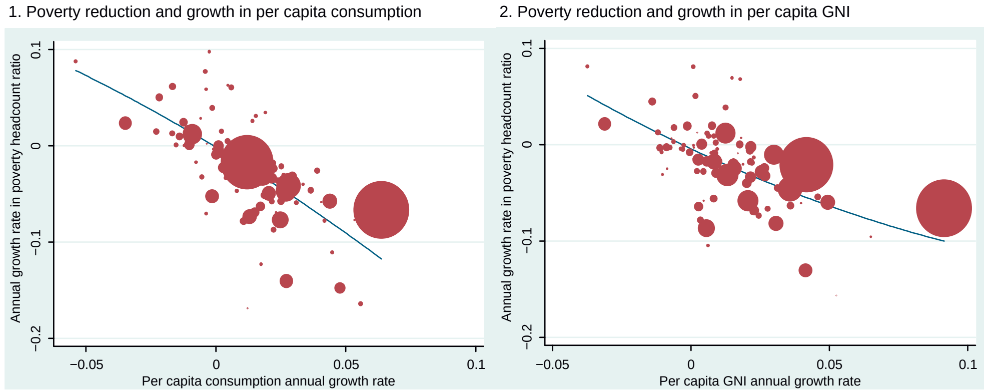


Figure 9.25 Growth and poverty reduction in the developing countries, 1981–2010. *Note: The horizontal axis shows the annual growth rate between 1981 and 2010 in per capita consumption from household surveys in constant 2005 dollars (panel 1) and in per capita gross national income in constant 2005 PPP dollars (panel 2); the vertical axis shows the annual growth change in the poverty headcount ratio ($1.25 line) in the same period. The size of each circle is proportional to the country population in 2010. Source: Own calculations based on PovcalNet (2013).*

Figure 9.25 is another illustration of the relationship between poverty reduction and economic growth. The figure shows that the change in poverty is closely negatively related to economic growth, either in per capita consumption expenditures drawn from household surveys or in GNI from NAS. Notice that in both panels the fitted line approximately crosses the (0,0) point.

Because the positive correlation between poverty reduction and growth is a well-established result, research in this area is mainly focused on estimating the magnitude of the corresponding elasticity, an issue that may have relevant policy implications. "If (the elasticity) is reasonably high, then poverty reduction strategies almost exclusively relying on economic growth are probably justified. If it is low, however, ambitious poverty reduction strategies might have to combine both economic growth and some redistribution" (Bourguignon, 2003). The growth elasticity of poverty reduction is typically estimated by regressing the annualized proportional change in a poverty indicator against the annualized growth rate of mean income (per capita income or consumption from surveys, or per capita GDP, GNI, or aggregate private consumption from NAS) in a sample of growth spells. In a regression without controls the resulting coefficient is the total elasticity, as opposed to a partial elasticity that can be obtained by holding inequality and other factors constant.

Based on a data set for 67 countries over the period from 1981 to 1994, and using the $1 line, Ravallion and Chen (1997) find a central estimate for the poverty-growth elasticity of −3.1. Values estimated by other authors are somewhat lower (in absolute terms): around −2 in World Bank (2000), −1.6 in Bourguignon (2003), −2.6 in Adams (2004), and −1.8

in Ferreira and Ravallion (2009). More recently, Ravallion (2012) reports an elasticity of −1.4 for the \$1.25 line. The elasticity is lower when using the growth rate of consumption per capita from NAS (−0.8), and also lower when using a higher poverty line. The confidence intervals around the regression coefficient are typically wide. Ravallion (2007) reports that the 95% confidence interval implies that for a poverty level of 40%, an annual growth rate of 2% is consistent with poverty reductions ranging from 1% to 7%.

Estimates of the total growth elasticity of poverty reduction over the period 1981–2010 obtained with the latest version of PovcalNet data are shown in Table 9.20.[123] All the elasticities are negative and significant at 1%. The growth elasticity of poverty reduction, as measured by the proportion of individuals below \$1.25 a day, is around −1.5. The results in Table 9.20 confirm that elasticities are lower in absolute value when considering a higher poverty line. The result of lower elasticities using GNI from National Accounts holds but only until the 2000s, suggesting a change in the relative trends between consumption reported in household surveys and output estimated in NAS. In general, the absolute value of the elasticities estimated with both sources have increased in the 2000s compared to previous decades, suggesting that growth was better translated to the poor in that period of falling inequality. For instance, although the elasticity using the \$1.25 line and consumption per capita was −1.53 for the period 1981–2010, it amounted to −1.83 when restricting the analysis to the 2000s.

So far, we have reported *total* elasticities, which can be seen as summary measures of the comovements of poverty and growth. The literature has tried to improve this characterization by adding other variables into the analysis, typically measures of inequality and development. The empirical evidence supports the intuition that higher inequality tends to reduce the absolute value of the elasticity, by shifting the gains from growth away from the poor (Kraay, 2006; Ravallion, 1997, 2001).[124] For example, based on a sample of 65 countries during 1981–2005, and using the \$1 poverty line, World Bank (2005) reported that the poverty-growth elasticity is highest among low-inequality countries (with a value of approximately −4.0 for countries with Ginis in the mid-20s) and lowest among high-inequality countries (close to −1.0 for countries with a Gini coefficient around 60). The change in inequality is also found as a significant direct determinant of the elasticity. For instance, Ravallion (2001) finds that the median rate of reduction in the poverty headcount ratio (\$1 line) among growing economies was 10% per year among countries with falling inequality, and 1% per year for those countries with rising inequality.

[123] The sample includes 725 spells in 76 countries for which consistent information for the whole period is available. Results do not significantly vary when we restrict the analysis to more recent periods, which allows a larger sample of countries.

[124] Although the intuition is compelling, the result is theoretically ambiguous. See Ravallion (2007) for a proof, and Bourguignon (2003) for a case in which assuming log-normality yields an unambiguous result.

Table 9.20 Total poverty−growth elasticities
Developing countries

Index	Poverty line	Consumption per capita			GNI per capita		
		1981–2010	1990–2010	1999–2010	1981–2010	1990–2010	1999–2010
Headcount	1.25	−1.53	−1.56	−1.83	−1.46	−1.51	−2.16
	2	−1.39	−1.41	−1.68	−1.17	−1.19	−1.91
	4	−1.30	−1.38	−1.10	−0.78	−0.81	−1.21
Poverty gap	1.25	−1.75	−1.77	−1.87	−1.44	−1.47	−2.09
	2	−1.50	−1.47	−1.85	−1.35	−1.30	−1.83
	4	−1.41	−1.47	−1.39	−0.95	−0.98	−1.54
Squared PG	1.25	−2.18	−2.25	−1.87	−1.50	−1.52	−1.87
	2	−1.60	−1.52	−1.86	−1.44	−1.29	−1.91
	4	−1.40	−1.43	−1.54	−1.04	−1.04	−1.71

Note: All elasticities are significant at 1% level.
Source: Own calculations based on PovcalNet (2013).

Using a sample of 114 growth spells from the mid-1980s to the mid-1990s in 26 developing countries Bourguignon (2003) finds an elasticity of −1.6 in the model without controls and −2 when controlling for the change in the Gini. The latter specification increases the R^2 from 0.266 to 0.419, suggesting that the heterogeneity in distributive changes is as much responsible for the variation in poverty changes across growth spells as the heterogeneity in the speed of growth. Bourguignon (2003) also finds positive and significant coefficients for the interaction terms between the growth rate and both (i) the initial level of inequality, and (ii) the level of development (proxied by the poverty line over mean income). He also reported negative coefficients for the interaction terms between the change in the Gini and those two variables. The first four columns in Table 9.21 in general confirm the results in Bourguignon (2003), using a larger data set spanning three decades.[125]

The results suggest the possibility of a double dividend from reducing inequality: first, given a growth rate, lower inequality is associated directly with less poverty, and second, less inequality means more poverty reaction to a given growth rate, that is an acceleration of poverty reduction for a given rate of economic growth. Ravallion (2007) illustrated the relevance of lower inequality for the perspectives of poverty reduction assuming a country with a poverty incidence rate of 40% and a growth rate of 2% per year: with an initial Gini coefficient of 30 it would take 11 years to reduce poverty by half, whereas with a Gini of 60 it would take 35 years.

A more recent study Ravallion (2012) finds that the (absolute) growth elasticity of poverty reduction tends to be lower in countries with a higher initial poverty rate (see also the two last columns in Table 9.21). Ravallion (2012) finds a large attenuating effect of a higher initial poverty rate on the elasticity: at an initial poverty rate of 10% the elasticity is −2.2, while it falls to −0.5 at a poverty rate of 80%. The results are robust to the inclusion of additional interaction effects with the initial Gini coefficient, the partial elasticity of poverty reduction holding the Lorenz curve constant, the primary school enrollment rate, life expectancy, the price of investment goods, and regional dummies.

Unfortunately, although the characterization of the growth elasticity of poverty reduction discussed in this section is useful, the literature is still far from being able to clearly identify the structural factors that are behind the differences in the elasticities, and therefore it is relatively silent about the specific policies that could foster a larger impact of economic growth on poverty.

[125] Instead, unlike Bourguignon (2003) we fail to find a coefficient close to −1 for the interaction between the growth rate and the theoretical value of the poverty-growth elasticity under the lognormal assumption.

Table 9.21 Regressions for annualized proportional change in the poverty headcount ratio
Developing countries, 1981–2010

	Unconditional	Conditional on: Inequality and development			Conditional on: Poverty	
	(1)	(2)	(3)	(4)	(5)	(6)
g = growth rate in per capita consumption	−1.391	−1.467	−2.879	−2.910	−2.111	−2.073
	(0.147)***	(0.144)***	(0.570)***	(0.590)***	(0.206)***	(0.188)***
dGini = change in Gini coefficient		1.488	1.566	3.195		
		(0.265)***	(0.279)***	(1.206)***		
g* initial Gini coeficient			0.022	0.023		
			(0.013)*	(0.014)*		
g* (poverty line/mean consumption)			0.903	0.879		
			(0.094)***	(0.101)***		
dGini* initial Gini coefficient				−0.019		
				(0.027)		
dGini* (poverty line/mean consumption)				−1.666		
				(0.345)***		
g* initial poverty rate					0.021	0.021
					(0.003)***	(0.003)***
Log initial poverty rate						−1.559
						(0.927)*
Intercept	0.150	−0.008	0.187	0.064	−0.255	4.108
	(0.225)	(0.203)	(0.154)	(0.208)	(0.122)**	(2.655)
Observations	725	695	695	695	725	725
R^2	0.414	0.502	0.583	0.610	0.508	0.516

Note: Sample of 76 countries with information for the period 1981–2010. Poverty indicator = headcount ratio, $2-a-day line. Growth rates computed from per capita consumption in household surveys.
*significant at 10%, **significant at 5%, ***significant at 1%.
Source: Own calculations based on PovcalNet (2013).

9.6.7 Relative Poverty

So far we have presented evidence on *absolute* poverty, an option that requires keeping the real value of the poverty line fixed over time. The alternative is *relative* poverty, where the line is adjusted to reflect changes in a general measure of economic well-being. Although the measurement of relative poverty has not been common in the developing world, sustained economic growth has triggered the update of absolute poverty lines in some countries (e.g., China and India) and foster discussions on relative poverty. The measurement of relative, instead of absolute, poverty can be justified on two grounds (Chen and Ravallion, 2008; Sen, 1983). First, the very concept of poverty may depend on social norms that are likely different across countries and over the development process. Second, even for a fixed norm, an absolute line in the space of welfare may require a relative line in the space of consumption if individual welfare also depends on relative consumption with respect to the rest of the society where the person lives.

The typical relative poverty line is set as a constant proportion of the mean of the distribution, implying that poverty does not fall when all incomes grow at the same rate. This result is considered problematic by many authors who prefer weak versions in which the cost of inclusion is not a constant proportion of mean income. For instance, following Atkinson and Bourguignon (2001), Ravallion and Chen (2011) propose a poverty line that is fixed up to a critical value of the mean, where it rises with elasticity less than one. In particular, Ravallion and Chen (2011) set the line for country i at time t as $z_{it} = \max[\$1.25, (\$1.25 + M_{it})/2]$, where M_{it} is the country and date specific mean. Figure 9.26 reveals an overall trend decline in the incidence of weakly relative poverty in the developing world. The fall has been relatively meager, so that the number of poor by this

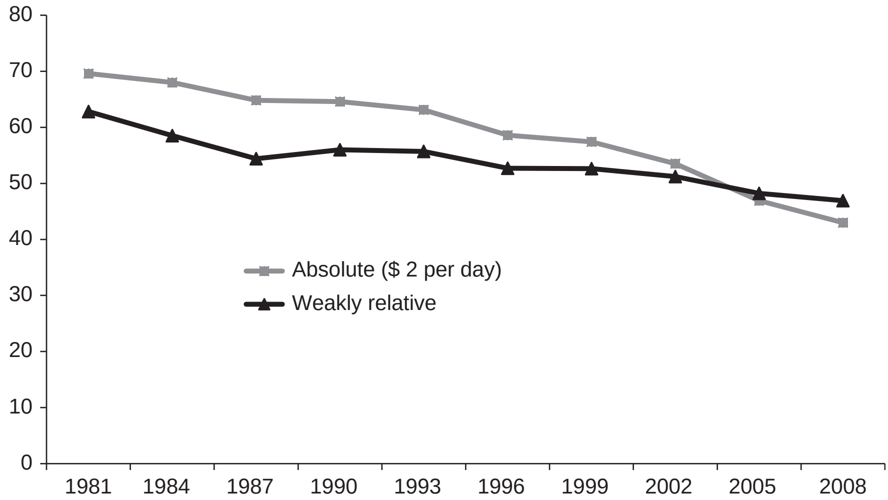


Figure 9.26 Poverty headcount ratio with absolute and weakly relative poverty lines. Developing countries, 1981–2008. *Source: Ravallion and Chen (2011).*

Table 9.22 Absolute and relative poverty
Developing countries, 1990–2010

	Unweighted				Weighted			
	absolute ($2)	weakly relative (RC)	strongly relative	anchored	absolute ($2)	weakly relative (RC)	strongly relative	anchored
1990	40.6	46.5	28.0	28.0	64.7	54.1	19.7	19.7
1993	39.8	47.9	29.0	32.9	62.7	53.7	21.1	19.1
1996	38.6	47.4	29.1	32.1	58.0	51.4	22.0	17.1
1999	38.3	47.5	30.1	31.9	56.7	50.2	22.9	16.9
2002	36.7	46.9	30.1	29.6	52.5	49.5	23.7	15.2
2005	32.3	45.0	29.5	24.8	45.6	46.5	23.7	12.3
2008	28.7	43.2	29.0	19.9	42.1	46.1	24.9	9.9
2010	27.7	42.4	28.6	18.9	39.4	45.6	26.6	9.5

Note: Weakly relative poverty computed according to Ravallion and Chen (2011). Strongly relative lines set at 50% of mean income. Anchored poverty computed by "anchoring" the relative lines (50% of mean income) to the values in 1990 and updating them only by domestic price changes.
Source: Own calculations based on data from PovcalNet (sample of 76 countries).

measure actually increased from 2.3 billion in 1981 to 2.7 billion in 2008. Although the coefficient in an OLS regression between the proportionate rate of poverty reduction (annualized differences in the logs) and the rate of growth in the mean is −1.89 (standard error = 0.23) for the absolute measure of poverty, it becomes −0.38 (standard error = 0.08) for the relative measure. Only East Asia has experienced a marked reduction in the incidence of relative poverty: from 81% in 1981 to 42% in 2008. In fact it is the only region in which the poverty count was reduced according to this measure. Weakly relative poverty measures rose or stayed unchanged in most regions in the 1990s and slowly went down in the 2000s.

In Table 9.22 we extend the estimates of Ravallion and Chen (2011) to 2010 and add estimates of strongly relative poverty (50% of mean income) and anchored poverty, computed by "anchoring" the relative lines (50% of mean income) to the values in 1990, updating them only by domestic price changes (see Chapter 8 in this volume). Our view of poverty trends widely varies across methodologies: Although the unweighted anchored poverty rate fell 32% between 1990 and 2010, the strongly relative poverty rate went up 2%. The differences are even larger when considering the population-weighted statistics.

9.7. CONCLUDING REMARKS

There has been a remarkable improvement in the availability of information for distributive analysis in the last decades due to increasing efforts by researchers, national governments, and international organizations. To be sure, the picture of inequality and poverty in the developing world is substantially sharper now than in the late 1990s, when the first volume of this Handbook was written. There remain, however, enormous data limitations that make that picture only a very rough approximation of reality. Household surveys are lacking in some countries and are carried out very occasionally in others. Changes in methodology over time are frequent, a fact that generally implies improvement in the data collection, but that at the same time introduces comparability issues with previous surveys that are difficult to deal with. Household surveys have endemic problems in capturing some income and consumption items and in dealing with selective compliance and underreporting issues. The gaps with National Accounts aggregates, usually variable over time, are a disturbing sign of measurement errors. Comparability across countries is another big problematic issue; there are few efforts among national agencies to standardize surveys or at least some criteria to gather and process information. Another issue of concern is the difficulty in obtaining statistical confidence intervals for the distributive statistics, either because agencies do not report them, or do not provide information on sampling issues, or do not release the microdata.[126] In sum, to be able to

[126] The methodology to compute statistical confidence intervals is easily implementable; see Chapter 6 of this Handbook.

characterize and track distributive changes with more accuracy, we need more efforts to extend the coverage and frequency of household surveys and improve their reliability and comparability across countries. There is still a long way to go to get an accurate assessment of poverty and inequality in the developing world.

From the data available, some general facts emerge. High poverty and inequality are pervasive characteristics of the developing world, but are not immutable features of these economies. The evidence suggests a robust decline in the levels of absolute income poverty in the developing world, driven mainly by East Asia in the 1980s and 1990s and generalized to the rest of the developing countries in the 2000s. Income poverty has been reduced in most countries and in the world as whole, making the achievement of the first MDG possible. Despite these positive results, there are reasons for concern. The task of fighting poverty continues to be very challenging: Around 1.2 billion people survive with less than \$1.25 a day, an extremely low standard. Also, most of the people who succeeded in jumping the \$1.25 line in the last decades are still poor by the standards of middle-income developing countries and remain highly vulnerable if economic conditions worsen. Moreover, due to economic growth, concerns about relative poverty may be increasingly important, and the evidence in that sense is less optimistic.

The evidence on relative income inequality suggests that on average the developing countries are somewhat more unequal than three decades ago. The patterns have been different by period and region. Inequality rose in the late 1980s and in the 1990s; the changes were larger in Eastern Europe and Central Asia, probably as a result of the transition from a central-planned to a more market-oriented economy; in East Asia, likely as a consequence of the economic takeoff; and in Latin America probably as the result of recurrent macroeconomic crises and some structural transformations. Distributive changes became more equalizing in the 2000s, but the changes were rather moderate and with considerable heterogeneity across countries. In fact, in this decade of widespread social improvement around a third of the countries did not experience falling inequality levels. Reducing inequality certainly remains a top concern in the developing world.

ACKNOWLEDGMENT

F. A. is with Nuffield College and INET-Oxford, Paris School of Economics and CONICET; L. G. is with Centro de Estudios Distributivos, Laborales y Sociales (CEDLAS), Facultad de Ciencias Económicas, Universidad Nacional de La Plata, and CONICET. This chapter was completed while L. G. was visiting professor at the University of British Columbia. We are grateful to participants to the conference "Recent Advances in the Economics of Income Distribution" (Paris), the AAEP Meetings (Rosario), Thomas Piketty, and especially Anthony Atkinson and Francois Bourguignon, for valuable comments and suggestions. We are also grateful to David Jaume, Darío Tortarolo, Carolina López, Julian Amendolaggine, Santiago Garganta, Florencia Pinto, Pablo Gluzmann, Leopoldo Tornarolli, Javier Alejo, Juan Zoloa, and Carolina García Domench (all at CEDLAS) for outstanding research assistance. We alone are responsible for any errors. Financial support from the ESRC-DFID joint fund and INET is gratefully acknowledged.

APPENDIX

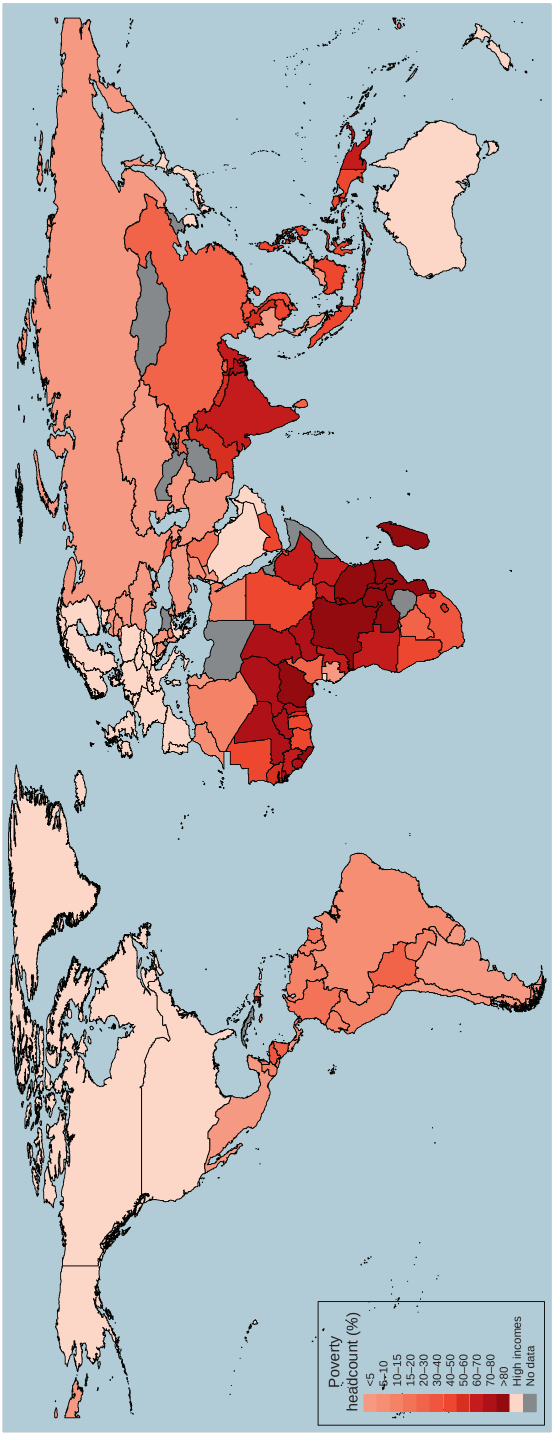


Figure A.1 Map of poverty in the developing world, circa 2010. *Note: Poverty computed over the distribution of consumption/income per capita with the PPP-adjusted \$2-a-day line. Source: Own calculations based on PovcalNet (2013).*

Table A.1 List of developing countries by region and population, 2010

Countries	Pop.	Data
East Asia and Pacific		
American Samoa	0.066	
Cambodia	13.823	X
China	1324.655	X
Fiji	0.844	X
Indonesia	234.951	X
Kiribati	0.097	
Korea, Dem. Rep.	24.126	
Lao	6.022	X
Malaysia	27.502	X
Marshall Islands	0.053	
Micronesia	0.110	X
Mongolia	2.667	X
Myanmar	47.250	
Palau	0.020	
Papua New Guinea	6.549	
Philippines	90.173	X
Samoa	0.182	
Solomon Islands	0.510	
Thailand	68.268	X
Timor-Leste	1.078	X
Tonga	0.103	
Tuvalu	0.010	
Vanuatu	0.228	
Vietnam	85.122	X
Eastern Europe and Central Asia		
Albania	3.181	X
Armenia	3.079	X
Azerbaijan	8.763	X
Belarus	9.602	X
Bosnia and Herzegovina	3.774	X
Bulgaria	7.623	X
Croatia	4.434	X
Czech Republic	10.424	X
Estonia	1.341	X
Georgia	4.384	X
Hungary	10.038	X
Kazakhstan	15.674	X
Kosovo	1.747	
Kyrgyz Republic	5.319	X
Latvia	2.266	X
Lithuania	3.358	X

Table A.1 List of developing countries by region and population, 2010—cont'd

Countries	Pop.	Data
Macedonia	2.053	X
Moldova	3.570	X
Montenegro	0.629	X
Poland	38.126	X
Romania	21.514	X
Russian Federation	141.950	X
Serbia	7.350	X
Slovak Republic	5.407	X
Slovenia	2.021	X
Tajikistan	6.691	X
Turkey	70.924	X
Turkmenistan	4.918	X
Ukraine	46.258	X
Uzbekistan	27.303	X
Latin America and the Caribbean		
Antigua and Barbuda	0.087	
Argentina	39.714	X
Belize	0.322	X
Bolivia	9.618	X
Brazil	191.543	X
Chile	16.796	X
Colombia	45.006	X
Costa Rica	4.522	X
Cuba	11.267	
Dominica	0.068	
Dominican Republic	9.665	X
Ecuador	14.057	X
El Salvador	6.130	X
Grenada	0.104	
Guatemala	13.691	X
Guyana	0.752	X
Haiti	9.736	X
Honduras	7.303	X
Jamaica	2.687	X
Mexico	110.627	X
Nicaragua	5.636	X
Panama	3.406	X
Paraguay	6.230	X
Peru	28.463	X
St. Kitts and Nevis	0.051	
St. Lucia	0.170	X
St. Vincent and G.	0.109	

Continued

Table A.1 List of developing countries by region and population, 2010—cont'd

Countries	Pop.	Data
Suriname	0.515	X
Trinidad and Tobago	1.331	X
Uruguay	3.334	X
Venezuela	27.935	X
Middle East and North Africa		
Algeria	34.428	X
Djibouti	0.856	X
Egypt	78.323	X
Iran	72.289	X
Iraq	30.178	X
Jordan	5.787	X
Lebanon	4.167	
Libya	6.150	
Morocco	31.321	X
Syria	19.638	X
Tunisia	10.329	X
West Bank and Gaza	3.937	X
Yemen	22.627	X
South Asia		
Afghanistan	32.518	
Bangladesh	145.478	X
Bhutan	0.701	X
India	1190.864	X
Maldives	0.308	X
Nepal	28.905	X
Pakistan	167.442	X
Sri Lanka	20.217	X
Sub-Saharan Africa		
Angola	18.038	X
Benin	8.356	X
Botswana	1.955	X
Burkina Faso	15.515	X
Burundi	7.943	X
Cameroon	18.759	X
Cape Verde	0.487	X
Central African R.	4.238	X
Chad	10.654	X
Comoros	0.697	X
Congo, Dem. R.	62.475	X
Congo, Rep.	3.836	X
Côte d'Ivoire	18.987	X

Table A.1 List of developing countries by region and population, 2010—cont'd

Countries	Pop.	Data
Eritrea	4.948	
Ethiopia	79.446	X
Gabon	1.450	X
Gambia	1.636	X
Ghana	23.264	X
Guinea	9.559	X
Guinea-Bissau	1.454	X
Kenya	38.455	X
Lesotho	2.127	X
Liberia	3.658	X
Madagascar	19.546	X
Malawi	14.005	X
Mali	14.460	X
Mauritania	3.295	X
Mauritius	1.269	
Mozambique	22.333	X
Namibia	2.200	X
Niger	14.450	X
Nigeria	150.666	X
Rwanda	10.004	X
São Tomé & P.	0.160	X
Senegal	11.787	X
Seychelles	0.087	X
Sierra Leone	5.612	X
Somalia	8.922	
South Africa	48.793	X
South Sudan	8.977	
Sudan	32.438	X
Swaziland	1.032	X
Tanzania	42.268	X
Togo	5.777	X
Uganda	31.339	X
Zambia	12.380	X
Zimbabwe	12.452	

Note: An "X" in the Data column marks that the country has distributive information in PovcalNet.
Source: Population (in millions) is taken from the United Nations Demographic Yearbook.

REFERENCES

Acar, S., Dogruel, F., 2012. Sources of inequality in selected MENA countries. Struct. Chang. Econ. Dyn. 23 (3), 276–285.

Adams, R., 2004. Economic growth, inequality and poverty: estimating the growth elasticity of poverty. World Dev. 32 (12), 1989–2014.

Adams Jr., R., Page, J., 2003. Poverty, inequality and growth in selected Middle East and North Africa countries, 1980–2000. World Dev. 31 (12), 2027–2048.

Ahluwalia, M., Cartner, N., Chenery, H., 1979. Growth and poverty in developing countries. J. Dev. Econ. 6 (3), 299–341.

Ahuja, V., Bidani, B., Ferreira, F., Walton, M., 1997. Everyone's miracle? Revisiting poverty and inequality in East Asia. Directions in Development. The World Bank, Washington, D.C.

Alatas, V., Bourguignon, F., 2000. The evolution of the distribution of income during Indonesian fast growth: 1980–1996. Princeton University, Princeton (unpublished).

Alejo, J., 2012. Educación y Desigualdad: una metodología de descomposición basada en dos interpretaciones de la ecuación de Mincer. Evidencia para Argentina. In: Anales de la XLVII Reunión Anual de la Asociación Argentina de Economía Política, Trelew, Argentina.

Alejo, J., Bérgolo, M., Carbajal, F., 2013. Las Transferencias Públicas y su impacto distributivo: La Experiencia de los Países del Cono Sur en la década de 2000. CEDLAS working paper 141 (February, 2013).

Altimir, O., 1979. La Dimensión de la pobreza en América Latina. Cuadernos de la CEPAL, vol. 27. United Nations, Santiago.

Altimir, O., 1996. Cambios de la desigualdad y la pobreza en la América Latina. Trimest. Econ. 61 (1), 85–133.

Alvaredo, F., Cano, L., Can personal tax records be used to study income concentration in Ecuador? (Forthcoming).

Alvaredo, F., 2011. A note on the relationship between top income shares and the gini coefficient. Econ. Lett. 110 (3), 274–277.

Alvaredo, F., Atkinson, A.B., 2010. Colonial Rule, Apartheid and Natural Resources: Top Incomes in South Africa 1903–2007, CEPR DP 8155; series updated in the World Top Incomes Database.

Alvaredo, F., Londoño, J.V., 2013. High Incomes and Personal Taxation in a Developing Economy: Colombia 1993-2013, Commitment to Equity-CEQ Working Paper n. 12.

Alvaredo, F., Piketty, T., 2014. Measuring Top Incomes and Inequality in the Middle East: Data Limitations and Illustration with the Case of Egypt. Working paper 832, The Economic Research Forum (ERF) 2014, Egypt.

Amiel, Y., Cowell, F., 1999. Thinking About Inequality: Personal Judgment and Income Distributions. Cambridge University Press, Cambridge.

Anand, S., Kanbur, R., 1993. The Kuznets process and the inequality-development relationship. J. Dev. Econ. 40, 25–52.

Anand, S., Segal, P., 2008. What do we know about global income inequality? J. Econ. Lit. 46 (1), 57–94.

Anderson, E., 2005. Openness and inequality in developing countries: a review of theory and recent evidence. World Dev. 33 (7), 1045–1063.

Aron, J., Kahn, B., Kingdon, G., 2009. South Africa Economic Policy under Democracy. Oxford University Press, Oxford.

Artadi, E., Sala-i-Martin, X., 2003. The Economic Tragedy of the XXth Century: Growth in Africa, NBER Working Paper 9865.

Asian Development Bank (ADB), 2012. Asian Development Outlook 2012. Confronting Rising Inequality in Asia. ADB, Mandaluyong City, Philippines.

Asian Development Bank, 2012a. Key Indicators for Asia and the Pacific 2012, 43rd ed. Asian Development Bank, Mandaluyong City, Philippines.

Asian Development Bank, 2012b. Asian Development Outlook 2012. Confronting Rising Inequality in Asia. ADB. Mandaluyong City, Philippines.

Atkinson, A., 1987. On the measurement of poverty. Econometrica 55 (4), 749–764.

Atkinson, A., Bourguignon, F., 2001. Poverty and inclusion from a world perspective. In: Stiglitz, J.E., Muet, P.A. (Eds.), Governance, Equity and Global Markets. Oxford University Press, Oxford.

Atkinson, A., Brandolini, A., 2001. Promise and pitfalls in the use of "secondary" data-sets: income inequality in OECD countries as a case study. J. Econ. Lit. 39 (3), 771–799.
Atkinson, A., Brandolini, A., 2004. Global world inequality: absolute, relative or intermediate? In: Paper Presented for the 28th Generation Conference of the International Association for Research in Income and Wealth, Cork, Ireland.
Atkinson, A., Brandolini, A., 2010. On analyzing the world distribution of income. World Bank Econ. Rev. 24 (1), 1–37.
Atkinson, A.B., Lugo, M.A., 2010. Growth, Poverty and Distribution in Tanzania, LSE Research Online Documents on Economics. London School of Economics and Political Science, LSE Library.
Atkinson, A., Marlier, E., 2010. Income and Living Conditions in Europe. Eurostat Statistical Books, Publication Office of European Union, Luxembourg.
Atkinson, A.B., Micklewright, J., 1992. Economic Transformation in Eastern Europe and the Distribution of Income. Cambridge University Press, Cambridge. Cambridge Books.
Atkinson, A.B., Morelli, S., 2011. Economic Crisis and Inequality, Human Development Research Paper 2011/06, UNDP.
Atkinson, A.B., Morelli, S., 2012. Chartbook of Economic Inequality: 25 Countries 1911-2010. No 15, INET Research Notes. Institute for New Economic Thinking (INET).
Atkinson, A.B., Piketty, T., 2007. Top Incomes Over the Twentieth Century: A Contrast Between Continental European and English-Speaking Countries. Oxford University Press, Oxford, UK.
Atkinson, A.B., Piketty, T. (Eds.), 2010. Top Incomes: A Global Perspective. Oxford University Press, Oxford.
Azevedo, J., Inchauste, G., Sanfelice, V., 2011. Decomposing the Recent Inequality Decline in Latin America. World Bank, manuscript.
Banerjee, A., Duflo, E., 2003. Inequality and growth: what can the data say? J. Econ. Growth 8 (3), 267–299.
Banerjee, A., Piketty, T., 2010. Top Indian incomes 1922-2000. In: Atkinson, A.B., Piketty, T. (Eds.), Top Incomes: A Global Perspective. Oxford University Press, Oxford, pp. 1–39, Chapter 1.
Barros, R.P., Carvalho, M., Franco, S., Mendoça, R., 2006. Uma análise das principais causas da queda recente na desigualdade de renda brasileira. Rev. Econ. 8 (1), 117–147.
Barros, R., de Carvalho, M., Franco, S., Mendonça, R., 2010. Markets, the state, and the dynamics of inequality in Brazil. In: López Calva, L., Lustig, N. (Eds.), Declining inequality in Latin America: A decade of progress?. Brookings Institution and UNDP.
Bastagli, F., Coady, D., Gupta, S., 2012. Income Inequality and Fiscal Policy, IMF Staff Discussion Note, SDN/12/08, June.
Bénabou, R., 1996. Inequality and growth. NBER macroeconomics annual 1996, vol. 11, NBER. (Chapter in NBER book NBER Macroeconomics Annual 1996, Volume 11 (1996), Ben S. Bernanke and Julio J. Rotemberg, Editors (p. 11–92), Conference held March 8–9, 1996, Published in January 1996 by MIT Press, in NBER Book Series NBER Macroeconomics Annual.).
Besley, T., Burgess, R., 2003. Halving global poverty. J. Econ. Perspect. 17 (3), 3–22.
Bhalla, S., 2002. Imagine There's No Country: Poverty, Inequality and Growth in the Era of Globalization. Institute for International Economics. Washington, D.C.
Bibi, S., Nabli, M., 2010. Equity and inequality in the Arab Region: Policy Research Report 33. Economic Research Forum. Cairo, Egypt.
Bigsten, A., Shimeles, A., 2003. Prospect for Pro-Poor Growth Strategies in Africa, WIDER Research Paper 42/2004.
Birdsall, N., Nellis, J., 2003. Winners and losers: assessing the distributional impact of privatization. World Dev. 31 (10), 1617–1633.
Bleaney, M., Nishiyama, A., 2003. Convergence in inequality: differences between advanced and developing countries. Econ. Bull. 4, 1–10.
Bourguignon, F., 2003. The growth elasticity of poverty reduction: explaining heterogeneity across countries and time periods. In: Eichler, T., Turnovsky, S. (Eds.), Inequality and Growth: Theory and Policy Implications. MIT Press, Cambridge, MA, pp. 2–26.
Bourguignon, F., Bussolo, M., 2012. Income distribution and computable general equilibrium models: macro-micro modelling. In: Dixon, Peter B., Jorgenson, Dale W. (Eds.), In: Handbook of Computable General Equilibrium Modeling, vol. 1A. North Holland Press.

Bourguignon, F., Morrison, C., 2002. Inequality among world citizens: 1820-1992. Am. Econ. Rev. 92 (4), 727–744.
Bourguignon, F., Levin, V., Rosenblatt, D., 2004. Declining international inequality and economic divergence: reviewing the evidence through different lenses. Écon. Int. 100, 13–25.
Bourguignon, F., Ferreira, F., Lustig, N., 2005. The Microeconomics of Income Distribution Dynamics in East Asia and Latin America. The World Bank, Washington, DC.
Bourguignon, F., Bussolo, M., Pereira da Silva, L., 2008a. The Impact of Macroeconomic Policies on Poverty and Income Distribution. Macro-Micro Evaluation Techniques and Tools. The World Bank, Washington, DC.
Bourguignon, F., Ferreira, F., Leite, P., 2008b. Beyond Oaxaca Blinder: accounting for differences in households income distributions. J. Econ. Inequal. 6 (2), 117–148.
Burkhauser, R., Feng, S., Jenkins, S., Larrimore, J., 2012. Recent trends in top income shares in the United States: reconciling estimates from March CPS and IRS Tax Return Data. Rev. Econ. Stat. 94 (2), 371–388.
Cameron, L., 2002. Growth with or without equity? The distributional impact of Indonesian development. Asia. Pac. Econ. Lit. 16, 1–17.
Caselli, F., Michaels, G., 2013. Do oil windfalls improve living standards? Evidence from Brazil. Am. Econ. J. 5 (1), 208–238.
Caselli, F., Esquivel, G., Lefort, F., 1996. Reopening the convergence debate: a new look at cross-country growth empirics. J. Econ. Growth 1, 363–389.
Chakravarty, S., 1987. Development Planning: The Indian Experience. Oxford University Press, New Delhi.
Chen, S., Ravallion, M., 2001. How did the world's poorest fare in the 1990s? Rev. Income Wealth 47 (3), 283–300.
Chen, S., Ravallion, M., 2008. China is Poorer than We Thought, but no Less Successful in the Fight against Poverty. World Bank Policy Research Working Paper No. 4621.
Chen, S., Ravallion, M., 2010. The developing world is poorer than we thought, but no less successful in the fight against poverty. Q. J. Econ. 125 (4), 1577–1625.
Chen, S., Ravallion, M., 2012. More Relatively-Poor People in a Less Absolutely-Poor World, The World Bank Policy Research Working Paper 6114.
Chotikapanich, D., Valenzuela, R., Prasada Rao, D., 1997. Global and regional inequality in the distribution of income: estimation with limited and incomplete data. Empir. Econ. 22 (4), 533–546.
Christiansen, L., Demery, L., Paternostro, S., 2002. Growth, Distribution and Poverty in Africa: Messages from the 1990s, The World Bank Policy Research Paper 2810.
Chusseau, N., Hellier, J., 2012. Globalization and Inequality: Where do We Stand? ECINEQ Working Paper 279.
Cockburn, J., Duclos, J., Zabsonré, A., 2012. Is the Value of Humanity Increasing? A Critical-Level Enquiry, FERDI Working Paper 52.
Conte, A., Vivarelli, M., 2007. Globalization and Employment: Imported Skill Biased Technological Change in Developing Countries, IZA Discussion Paper 2797.
Cornia, G., 1996. Transition and Income Distribution: Theory, Evidence and Initial Interpretation, UNU-WIDER Research in Progress 1.
Cornia, G., 2011. Economic integration, inequality and growth: Latin America vs. the European economies in transition. Rev. Econ. Inst. 2 (2), 1–31
Cornia, G., Kiiski, S., 2001. Trends in Income Distribution in the Post-World War II Period: Evidence and Interpretation, UNU–WIDER Discussion Paper, 2001/89.
Cornia, G., Reddy, S., 2001. The Impact of Adjustment Related Social Funds on Income Distribution and Poverty, UNU/WIDER Discussion Paper 2001/1.
Cruces, G., Gasparini, L. 2012. Políticas sociales para la reducción de la desigualdad y la pobreza en América Latina y el Caribe. Diagnóstico, propuesta y proyecciones en base a la experiencia reciente. Premio Fundación Vidanta 2012, Contribuciones a la reducción de la pobreza y la desigualdad en América Latina y el Caribe, Fundación Vidanta, México.
Cruces, G., García Domench, C., Gasparini, L., 2014. Inequality in education: evidence for Latin America. In: Cornia, G. (Ed.), Falling Inequality in Latin America. Policy Changes and Lessons. Oxford University Press, pp. 318–339.

Datt, G., Ravallion, M., 1992. Growth and redistribution components of changes in poverty measures: a decomposition with applications to Brazil and India in the 1980s. J. Dev. Econ. 38 (2), 275–295.
Deaton, A., 1997. The Analysis of Household Surveys: A Microeconometric Approach to Development Policy. The World Bank, Washington, DC.
Deaton, A., 2003. How to monitor poverty for the millennium development goals. J. Hum. Dev. 4 (3), 353–378.
Deaton, A., 2005. Measuring poverty in a growing world (or measuring growth in a poor world). Rev. Econ. Stat. 87 (1), 1–19.
Deaton, A., 2010. Price indexes, inequality, and the measurement of world poverty. Am. Econ. Rev. 100 (1), 5–34.
Deaton, A., Dupriez, O., 2008. Poverty PPPs Around the World: An Update and Progress Report. Development Data Group, World Bank, Washington, D.C.
Deaton, A., Heston, A., 2010. Understanding PPPs and PPP-based national accounts. Macroeconomics 2 (4), 1–35.
Deaton, A., Zaidi, S., 2002. Guidelines for Constructing Consumption Aggregates for Welfare Analysis. The World Bank, Living Standards Measurement Study Working Paper 135.
De Ferranti, D., Perry, G., Ferreira, F., Walton, M., 2004. Inequality in Latin America, Breaking with History? The World Bank, Washington, DC.
Deininger, K., Squire, L., 1996. A new data set measuring income inequality. World Bank Econ. Rev. 10 (3), 565–591.
Devarajan, S., Nabi, I., 2006. Economic Growth in South Asia: Promising, Un-Equalizing,...Sustainable? The World Bank South Asia Region, Washington, D.C.
Dhongde, S., Miao, X., 2013. Cross-Country Convergence in Income Inequality, ECINEQ Working Papers 290.
Dhongde, S., Minoui, C., 2010. Global Poverty Estimates: Present and Future, ECINEQ Working Papers 181.
Dollar, D., Kraay, A., 2002. Growth is good for the poor. J. Econ. Growth 7 (3), 195–225.
Dollery, B., 2003. A history of inequality in South Africa, 1652–2002. South African Journal of Economics, Economic Society of South Africa 71 (3), 595–610, 09.
Essama-Nssah, B., 2012. Identification of Sources of Variation in Poverty Outcomes, The World Bank Policy Research Working Papers 5954.
Feenstra, R., 2008. Offshoring in the global economy: lecture 1: microeconomic structure; lecture 2: macroeconomic implications. In: The Ohlin Lectures, presented at the Stockholm School of Economics on September 17-18, 2008.
Ferreira, F., 1999. Economic transition and the distributions of income and wealth. Econ. Transit. 7 (2), 377–410.
Ferreira, F., 2010. Distributions in Motion: Economic Growth, Inequality, and Poverty Dynamics, The World Bank Policy Research Working Paper 5424.
Ferreira, F.H.G., Paes de Barros, R., 1999. The slippery slope: explaining the increase in extreme poverty in Urban Brazil, 1976–1996. Brazilian Review of Econometrics 19 (2), 211–296.
Ferreira, F.H.G., Leite, P.G., Litchfield, J.A., 2006. The Rise and Fall of Brazilian Inequality: 1981–2004. World Bank Policy Research Working Paper 3867. World Bank, Washington, DC.
Ferreira, F.H.G., Leite, P.G., Wai-Poi, M., 2007. Trade Liberalization, Employment Flows and Wage Inequality in Brazil. Research Paper, UNUWIDER, United Nations University (UNU), No. 2007/58.
Ferreira, F., Ravallion, M., 2009. Poverty and inequality: the global context. The Oxford Handbook of Economic Inequality. Oxford University Press, Oxford.
Ferreira, F., Leite, P., Wai-Poi, M., 2010. Trade liberalization, employment flows, and wage inequality in Brazil. In: Nissanke, M., Thorbecke, E. (Eds.), The Poor Under Globalization in Asia, Latin America and Africa. Oxford University Press, Oxford.
Fields, G., 2002. Distribution and Development: A New Look at the Developing World. MIT Press, Cambridge.
Fiszbein, A., Schady, N., 2009. Conditional cash transfers. Reducing present and future poverty: The World Bank Policy Research Report.

Fleisher, B.M., Sabirianova, K., Wang, X., 2005. Returns to skills and the speed of reforms: evidence from Central and Eastern Europe, China, and Russia. J. Comp. Econ. 33 (2), 351–370.
Foguel, M., Azevedo, J.P., 2007. Uma decomposição da desigualdade de rendimentos no Brasil: 1984–2005. In: Barros, R., Foguel, M., Ulyssea, G. (Eds.), Desigualdade de Renda no Brasil: uma análise da queda recente. IPEA, Brasília.
Forbes, K., 2011. Inequality in Crisis and Recovery: Revealing the Divides: The Case of Brazil. International Labour Organization, ILO Working Papers 469849.
Fortin, N., Lemieux, T., Firpo, S., 2011. Decomposition methods in economics. In: Ashenfelter, O., Card, D. (Eds.), In: Handbook of Labor Economics, vol. 4, Part A.
Fosu, A., 2009. Inequality and the impact of growth on poverty: comparative evidence for Sub-Saharan Africa. J. Dev. Stud. 45 (5), 726–745.
Fosu, A., 2010. The Global Financial Crisis and Development: Whither Africa? UNU-WIDER Working Paper 2010/124.
Fosu, A., 2011. Growth, Inequality, and Poverty Reduction in Developing Countries. Recent Global Evidence, UNU-WIDER Working Paper 2011/1.
Galbraith, J., Kum, H., 2005. Estimating the inequality of household incomes: toward a dense and consistent global data set. Rev. Income Wealth 51 (1), 115–143.
Garcia, M., Moore, C., 2012. The Cash Dividend: The Rise of Cash Transfer Programs in Sub-Saharan Africa. The World Bank, Washington, DC.
Gasparini, L., Gluzmann, P., 2012. Estimating income poverty and inequality from the Gallup World Poll: the case of Latin America and the Caribbean. J. Income Distrib. 21 (1), 3–27
Gasparini, L., Lustig, N., 2011. The rise and fall of income inequality in Latin America. In: Ocampo, J.A., Ros, J. (Eds.), The Oxford Handbook of Latin American Economics. Oxford University Press, Oxford, Chapter 27.
Gasparini, L., Cruces, G., Tornarolli, L., 2011a. Recent trends in income inequality in Latin America. Economia 10 (2), 147–201.
Gasparini, L., Galiani, S., Cruces, G., Acosta, P., 2011b. Educational Upgrading and Returns to Skills in Latin America: Evidence from a Supply-Demand Framework, 1990-2010, IZA Discussion paper 6244.
Gasparini, L., Cicowiez, M., Sosa Escudero, W., 2013. Pobreza y desigualdad en América Latina. Conceptos, herramientas y aplicaciones. Editorial Temas. Buenos Aires, Argentina.
Gentilini, U., Sumner, A., 2012. Poverty Where People Live: What do National Poverty Lines Tell Us About Global Poverty? International Policy Centre for Inclusive Growth Working Paper 98.
Gertler, P., 2004. Do conditional cash transfers improve child health? Evidence from PROGRESA's control randomized experiment. Am. Econ. Rev. 94 (2), 336–341.
Ghani, E., 2010. The Poor Half Billion in South Asia. What is Holding Back Lagging Regions? Oxford University Press, Published in India by Oxford University Press, New Delhi.
Gluzmann, P., 2012. Desigualdad del ingreso y del bienestar subjetivo. Análisis y comparaciones internacionales. PhD dissertation, Universidad Nacional de La Plata.
Go, D., Nikitin, D., Wang, X., Zou, H., 2007. Poverty and inequality in Sub-Saharan Africa: literature survey and empirical assessment. Ann. Econ. Fin. Soc. 8 (2), 251–304.
Goñi, E., López, H., Servén, L., 2008. Fiscal Redistribution and Income inequality in Latin America. Policy Research Working Paper 4487, World Bank.
Goldberg, P., Pavcnik, N., 2007. Distributional effects of globalization in developing countries. J. Econ. Lit. 45 (1), 39–82.
Gorodnichenko, Y., Sabirianova, K., 2005. Returns to schooling in Russia and Ukraine: a semiparametric approach to cross-country comparative analysis. J. Comp. Econ. 33 (2), 324–350.
Gosh, J., 2012. Unequal in Asia. Frontline 29 (15), 127–128, Jul 28-Aug 10.
Harrison, A., McLaren, J., McMillan, M., 2011. Recent perspectives on trade and inequality. Annu. Rev. Econom. 3 (1), 261–289.
Hellier, J., Lambrecht, S., 2012. Inequality, Growth and Welfare: The Main Links, ECINEQ Working Paper 258.
Helpman, E., Trajtenberg, M., 1998. A time to sow and a time to reap: growth based on general purpose technologies. In: Helpman, E. (Ed.), General Purpose Technologies and Economic Growth. MIT Press, Cambridge.

Heston, A., Summers, R., Aten, B., 2012. Penn World Table Version 7.1, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, July.
Hoffmann, R., 2006. Queda da desigualdade da distribuição de renda no Brasil, de 1995 a 2005, e delimitação dos relativamente ricos em 2005. In: Barros, R., Foguel, M., Ulyssea, G. (Eds.), Desigualdade de Renda no Brasil: uma análise da queda recente. IPEA, Brasília.
Hussain, A., 2008. Power Dynamics, Institutional Instability and Economic Growth: The Case of Pakistan. The Asia Foundation.
IDB, 1999. América Latina frente a la desigualdad. Informe de Progreso Económico y Social, Banco Interamericano de Desarrollo.
ILO, 2010. World Social Security Report 2010/11: Providing Coverage in Times of Crisis and Beyond. International Labour Office, Geneva.
Inchauste, G., Olivieri, S., Saavedra, J., Winkler, H., 2012. What is Behind the Decline in Poverty Since 2000? Evidence from Bangladesh, Peru and Thailand, The World Bank Policy Research Working Paper 6199.
Iqbal, F., 2006. Sustaining Gains in Poverty Reduction and Human Development in the Middle East and North Africa. The World Bank, Washington, DC.
Ivaschenko, O., 2002. Growth and Inequality: Evidence from Transitional Economies, CESIFO Working Paper 746.
Jain, S., 1975. Size Distribution of Income. A Compilation of Data. The World Bank, Washington, DC.
Jomo, K.S., Baudot, J., 2007. Flat World, Big Gaps. Economic Liberalization, Globalization and Inequality. Zed Books, London.
Kakwani, N., 2000. On measuring growth and inequality components of poverty with application to Thailand. J. Quant. Econ. 16, 67–68.
Kakwani, N., Subbarao, K., 1990. Rural poverty and its alleviation in India. Econ. Polit. Wkly. 25 (3), 2–16
Kanbur, R., Sumner, A., 2011. Poor Countries or Poor People? Development Assistance and the New Geography of Global Poverty, Cornell University Working Paper 2011-08.
Karshenas, M., 2003. Global poverty: national accounts based versus survey based estimates. Dev. Chang. 34 (4), 683–712.
Khan, A.R., 2008. Measuring inequality and poverty in Bangladesh: an assessment of the survey data. Bangl. Dev. Stud. 31 (3/4), 1–34.
Klasen, S., 1997. Poverty, inequality and deprivation in South Africa: an analysis of the 1993 SALDRU survey. Soc. Indic. Res. 41 (1–3), 51–94.
Klasen, S., 2005. Measuring poverty and deprivation in South Africa. Rev. Income Wealth 46 (1), 33–58.
Korinek, A., Mistiaen, J., Ravallion, M., 2006. Survey nonresponse and the distribution of income. J. Econ. Inequal. 4 (2), 33–55
Kraay, A., 2006. When is growth pro-poor? Evidence from a panel of countries. J. Dev. Econ. 80 (1), 198–227.
Kuznets, S., 1955. Economic growth and income inequality. Am. Econ. Rev. 45 (1), 1–28.
Langoni, C.G., 2005. Distribuição de Renda e Desenvolvimento Econômico no Brasil. Editora FGV, Rio de Janeiro.
Leibbrandt, M., Woolard, I., Woolard, C., 2009. A long-run perspective on contemporary poverty and inequality dynamics. In: Aron, J., Kahn, B., Kingdon, G. (Eds.), South Africa Economic Policy Under Democracy. Oxford University Press, Oxford, Chapter 10.
Leibbrandt, M., Woolard, I., McEwen, H., Koep, C., 2010a. Better employment to reduce inequality further in South Africa. In: OECD, Tackling Inequalities in Brazil, China, India and South Africa: The Role of Labour Market and Social Policies. OECD Publishing.
Leibbrandt, M., Woolard, I., Finn, A., Argent, J., 2010b. Trends in South African Income Distribution and Poverty since the Fall of Apartheid, OECD Social, Employment and Migration Working Papers n. 101.
Leigh, A., 2007. How closely do top income shares track other measures of inequality? Econ. J. 117, F619–F633.
Lewis, A., 1954. Economic development with unlimited supplies of labor. Manch. Sch. Econ. Soc. Stud. 22, 139–192.

Li, H., Squire, L., Zou, H., 1998. Explaining international and intertemporal variations in income inequality. Econ. J. 108 (446), 26–43.
Li, S., Chuliang, L., Sicular, T., 2013. Overview: income inequality and poverty in China, 2002–2007. In: Shi, L., Sato, H., Sicular, T. (Eds.), Rising Inequality in China: Challenge to the Harmonious Society. Cambridge University Press, New York.
Lind, J., Mehlum, H., 2010. With or without U? The appropriate test for a U shaped relationship. Oxf. Bull. Econ. Stat. 72 (1), 109–118.
Loayza, N., Raddatz, C., 2010. The composition of growth matters for poverty alleviation. J. Dev. Econ. 93 (1), 137–151.
Londoño, J., Székely, M., 2000. Persistent poverty and excess inequality: Latin America, 1970-1995. J. Appl. Econ. 3 (1), 93–134.
López Calva, L., Lustig, N. (Eds.), 2010. Declining Iinequality in Latin America: A Decade of Progress? Brookings Institution and UNDP.
Lustig, N., 1990. Economic crisis, adjustment and living standards in Mexico, 1982-85. World Dev. 18 (10), 1325–1342.
Lustig, N., 1995. Coping with Austerity: Poverty and Inequality in Latin America. Brookings Institution, Washington, DC.
Lustig, N., 2012. Fiscal Policy and Income Redistribution in Latin America: Challenging the Conventional Wisdom, Tulane University Economics Working Paper 1124.
Lustig, N., Lopez-Calva, F., Ortiz-Juarez, E., 2012. Declining Inequality in Latin America in the 2000s: The Cases of Argentina, Brazil, and Mexico. Working paper 307, Center for Global Development.
Mahmoudi, V., 2001. Growth-Equity Decomposition of a Change in Poverty: An Application to Iran. University of Essex, Colchester, UK.
Mazundar, D., 2012. Decreasing poverty and increasing inequality in India. In: OECD, (Ed.), Tackling Inequalities in Brazil, China, India and South Africa: The Role of Labour Market and Social Policies. OECD Publishing, pp. 157–207, Chapter 4.
McGrath, M.D., 1983. The Distribution of Personal Income in South Africa in Selected Years over the Period from 1945 to 1980. PhD thesis, University of Natal, Durban.
McGrath, M.D., Whiteford, A., 1994. The Distribution of Income in South Africa. Human Science Research Council, Pretoria.
Milanovic, B., 1998. Income, Inequality, and Poverty During the Transition from Planned to Market Economy. World Bank Regional and Sectoral Studies, The World Bank, Washington, DC.
Milanovic, B., 1999. Explaining the increase in inequality during transition. Econ. Transit. 7 (2), 299–341.
Milanovic, B., 2002. True world income distribution, 1988 and 1993: first calculation based on household surveys alone. Econ. J. 112 (476), 51–92.
Milanovic, B., 2003. Is Inequality in Africa Really Different? The World Bank Policy Research Working Paper 3169.
Milanovic, B., 2005. Worlds Apart: Measuring International and Global Inequality. Princeton University Press, Princeton.
Milanovic, B., 2012. Global Income Inequality by the Numbers: History and Now. An Overview, The World Bank Policy Research Working Paper 6259.
Milanovic, B., Ersado, L., 2010. Reform and Inequality During the Transition. An Analysis using Panel Household Survey Data, 1990-2005. UNU/WIDER Working Paper 62.
Minoiu, C., Reddy, S.G., 2008. Chinese poverty: assessing the impact of alternative assumptions. Rev. Income Wealth 54 (4), 572–596.
Miranti, R., Vidyattama, Y., Hansnata, E., Cassells, R., Duncan, A., 2013. Trends in Poverty and Inequality in Decentralising Indonesia, OECD Social, Employment and Migration Working Papers No. 148.
Mitra, P., Yemtsov, R., 2006. Increasing Inequality in Transition Economies: Is There More to Come? The World Bank Policy Research Working Paper 4007.
Moradi, A., Baten, J., 2005. Inequality in Sub-Saharan Africa: new data and new insights from anthropometric estimates. World Dev. 33 (8), 1233–1265.
Mosberger, P., 2014. Capital and Labor: Top Income Shares in Hungary, 1915-2010. PhD thesis.

Nattrass, N., Seekings, J., 1997. Citizenship and welfare in South Africa: deracialisation and inequality in a labour-surplus economy. Can. J. Afr. Stud. 31 (3), 452–481.
Nissanke, M., Thorbecke, E. (Eds.), 2010. The Poor Under Globalization in Asia, Latin America and Africa. Oxford University Press, Oxford, UK.
Okojie, C., Shimeles, A., 2006. Inequality in Sub-Saharan Africa. The Inter-Regional Inequality Facility. DFID, London.
Page, J., 2007. Boom, bust, and the poor: poverty dynamics in the Middle East and North Africa, 1970–1999. Q. Rev. Econ. Finance 46 (5), 832–851.
Palma, J., 2011. Homogeneous Middles vs. Heterogeneous Tails, and the End of the "Inverted-U": the Share of the Rich is What it´s all About, Cambridge Working Papers in Economics 1111.
Parker, S., Skoufias, E., 2001. Conditional cash transfers and their impact on child work and school enrollment: evidence from the PROGRESA program in Mexico. Economia 2 (1), 45–96
Paukert, F., 1973. Income distribution at different levels of development: a survey of evidence. Int. Labour Rev. 108, 97–125.
Piketty, T., Qian, N., 2010. Income inequality and progressive income taxation in China and India, 1986-2015. In: Atkinson, A.B., Piketty, T. (Eds.), Top Incomes: A Global Perspective. Oxford University Press, Oxford and New York.
Pinkovskiy, M., 2013. World welfare is rising: estimation using nonparametric bounds on welfare measures. J. Public Econ. 97 (1), 176–195.
Pinkovskiy, M., Sala-i-Martin, X., 2009. Parametric Estimations of the World Distribution of Income, NBER Working Paper 1543.
Pissarides, C., 1997. Learning by trading and the returns to human capital in developing countries. World Bank Econ. Rev. 11 (1), 17–32.
PovcalNet, 2013. PovcalNet: An Online Poverty Analysis Tool. The World Bank. http://go.worldbank.org/WE8P1I8250.
Rama, M., 2003. Globalization and Workers in Developing Countries, The World Bank Policy Research Working Paper 2958.
Ravallion, M., 1997. Can high-inequality developing countries escape absolute poverty? Econ. Lett. 56 (1), 51–57.
Ravallion, M., 2001. Growth, inequality and poverty: looking beyond averages. World Dev. 29 (11), 1803–1815.
Ravallion, M., 2003. Inequality convergence. Econ. Lett. 80 (3), 351–356.
Ravallion, M., 2004. Competing Concepts of Inequality in the Globalization Debate, The World Bank Policy Research Working Paper 3243.
Ravallion, M., 2007. Inequality is bad for the poor. In: Jenkins, S., Micklewright, J. (Eds.), Inequality and Poverty Re-Examined. Oxford University Press, Oxford.
Ravallion, M., 2010. Price Levels and Economic Growth: Making Sense of the PPP Changes Between ICP Rounds, The World Bank Policy Research Working Paper 5229.
Ravallion, M., 2012. Why don't we see poverty convergence? Am. Econ. Rev. 102 (1), 504–523.
Ravallion, M., Chen, S., 1997. What can new survey data tell us about recent changes in distribution and poverty? World Bank Econ. Rev. 11 (2), 357–382.
Ravallion, M., Chen, S., 2007. China's (uneven) progress against poverty. J. Dev. Econ. 82 (1), 1–42.
Ravallion, M., Chen, S., 2011. Weakly relative poverty. Rev. Econ. Stat. 93 (4), 1251–1261.
Ravallion, M., Datt, G., van de Walle, D., 1991. Quantifying absolute poverty in the developing world. Rev. Income Wealth 37 (4), 345–361.
Ravallion, M., Chen, S., Sangraula, P., 2009. Dollar a day revisited. World Bank Econ. Rev. 23 (2), 163–184.
Reddy, S., Pogge, T., 2010. How not to count the poor. In: Anand, S., Segal, P., Stiglitz, J. (Eds.), Debates on the Measurement of Poverty. Oxford University Press, Oxford.
Robinson, J., 2009. The Political Economy of Inequality, ERF Working Paper 493.
Sala-i-Martin, X., 2006. The world distribution of income: falling poverty and . . . convergence, period. Q. J. Econ. 121 (2), 351–397.
Sala-i-Martin, X., Pinkovskiy, M., 2010. African Poverty is Falling . . . Much Faster than You Think! NBER Working Paper 15775.

Salehi-Isfahani, D., 2010. Human Development in the Middle East and North Africa, UNDP Human Development Research Paper 2010/26.
SEDLAC, 2013. Socio-Economic Database for Latin America and the Caribbean. CEDLAS and the World Bank. http://sedlac.econo.unlp.edu.ar/eng/index.php.
Sen, A., 1976. Real national income. Rev. Econ. Stud. 43 (1), 19–39.
Sen, A., 1983. Poor, relatively speaking. Oxf. Econ. Pap. 35 (2), 153–169.
Sharma, M., Inchauste, G., Feng, J., 2011. Rising Inequality with High Growth and Falling Poverty. An Eye on East Asia and Pacific, vol. 9, The World Bank, Washington, D.C.
Shorrocks, A., Wan, G., 2008. Ungrouping income distributions: synthesising samples for inequality and poverty analysis. In: Basu, K., Kanbur, R. (Eds.), The Oxford Handbook of Arguments for a Better World: Essays in Honor of Amartya Sen. In: Ethics, Welfare and Measurement, vol. I. Oxford University Press, Oxford.
Simai, M., 2006. Poverty and Inequality in Eastern Europe and the CIS Transition Economies, DESA Working Paper No. 17.
Simkins, C., 1991. The Urban Foundation Income Distribution Model: Phase One. Urban Foundation, Johannesburg, Mimeo.
Skoufias, E., Suryahadi, A., 2000. Changes in Regional Inequality and Social Welfare in Indonesia Between 1996 and 1999, Working Paper, Social Monitoring and Early Response Unit (SMERU) Research Institute, Jakarta.
Soares, S., Guerreiro Osório, R., Veras Soares, F., Medeiros, M., Zepeda, E., 2009. Conditional cash transfers in Brazil, Chile and Mexico: impacts upon inequality. Estudios Económicos, número extraordinario 0 (Special issue), 207–224.
Solt, F., 2009. Standardizing the world income inequality database. Soc. Sci. Q. 90 (2), 231–242.
Standing, G., Vaughan-Whitehead, D., 1995. Minimum Wages in Central and Eastern Europe: From Protection to Destitution. International Labour Office. Central and Eastern European Team, Budapest, Central European University Press.
Sumner, A., 2012. From Deprivation to Distribution: Is Global Poverty Becoming a Matter of National Inequality? IDS Working Paper 2012-394.
Székely, M., Hilgert, M., 1999. What's Behind the Inequality We Measure: An Investigation using Latin American Data, Inter-American Development Bank, Research Department Working Paper.
Terreblanche, S., 2002. A History of Inequality in South Africa, 1652-2002. University of Natal Press, Pietermaritzburg.
Tinbergen, J., 1975. Income Differences: Recent Research. North Holland Press, Amsterdam.
Todd, P., Wolpin, K., 2006. Assessing the impact of a school subsidy program in Mexico: using a social experiment to validate a dynamic behavioral model of child schooling and fertility. Am. Econ. Assoc. 96 (5), 1384–1417.
van der Berg, S., Louw, M., 2004. Changing patterns of South African income distribution: towards time series estimates of distribution and poverty. S. Afr. J. Econ. 72 (3), 546–572.
Van de Walle, N., 2008. The Institutional Origins of Inequality in Sub-Saharan Africa, Cornell University Mario Einaudi Center for International Studies Working Paper 6-08.
Veras Soares, F., Perez Ribas, R., Guerreiro Osório, R., 2007. Evaluating the Impact of Brazil's Bolsa Família: Cash Transfer Programmes in Comparative Perspective. International Poverty Centre, IPC Evaluation Note 1, 2007.
Voitchovsky, S., 2009. Inequality, growth and sectoral change. In: Salverda, W., Nolan, B., Smeeding, T.M. (Eds.), Oxford Handbook of Economic Inequality. Chapter 22.
Whiteford, A., Van Seventer, D., 2000. South Africa's changing income distribution in the 1990s. Stud. Econ. Econ. 24 (3), 7–30.
WIDER, 2008. World Income Inequality Database. User Guide and Data Sources. UNU-WIDER, Helsinki.
Winters, A., McCulloch, N., McKay, A., 2004. Trade liberalization and poverty: the evidence so far. J. Econ. Lit. 42, 72–115.
Wodon, Q., Castro, R., Lee, K., 2001. Poverty in Latin America: trends (1986-1998) and determinants. Cuad. Econ. 38 (114).

Wood, A., 1997. Openness and wage inequality in developing countries: the Latin American challenge to East Asian conventional wisdom. World Bank Econ. Rev. 11 (1), 33–57.
World Bank, 1990. World Development Report 1990: Poverty. The World Bank, Washington D.C.
World Bank, 2000. World Development Report: Attacking Poverty. Oxford University Press, New York.
World Bank, 2005. ECAPOV II. Growth, Poverty and Inequality in Eastern Europe and Former Soviet Union. The World Bank, Washington, DC.
World Bank, 2006a. World Development Report: Equity and Development. World Bank and Oxford University press. Washington, D.C.
World Bank, 2006b. Can South Asia Reduce Poverty in a Generation? http://go.worldbank.org/NF9FFJTM20.
World Bank, 2008. Global purchasing power parities and real expenditures. In: 2005 International Comparison Program. The World Bank, Washington, DC.
World Bank, 2009. The World Development Report 2009: Reshaping Economic Geography. The World Bank, Washington, DC.
Zin, R.H.M., 2005. Income distribution in East Asian developing countries: recent trends. Asia. Pac. Econ. Lit. 19 (2), 36–54.

CHAPTER 10

Income Mobility

Markus Jäntti*, **Stephen P. Jenkins**[†,‡,§]
*Swedish Institute for Social Research (SOFI), Stockholm University, Stockholm, Sweden
[†]Department of Social Policy, London School of Economics, London, UK
[‡]ISER, University of Essex, Colchester, UK
[§]IZA, Bonn, Bonn, Germany

Contents

Abstract

We survey the literature on income mobility, aiming to provide an integrated discussion of mobility within and between generations. We review mobility concepts, descriptive devices, measurement methods, data sources, and recent empirical evidence.

Keywords

Intragenerational mobility, Intergenerational mobility, Income mobility, Earnings mobility

Handbook of Income Distribution, Volume 2A
ISSN 1574-0056, http://dx.doi.org/10.1016/B978-0-444-59428-0.00011-4

JEL Classification Codes

D31, I30

10.1. INTRODUCTION

Most of the information that we have about income distribution is cross-sectional in nature; there are statistics about, for example, income levels, poverty rates, and the extent of inequality for a given year or for a series of years. The data sources used to provide estimates for the different years refer to different samples of individuals. In this chapter, we discuss a different but complementary perspective on income distribution to the cross-sectional one. We take an explicitly longitudinal perspective, one that is based on tracking over time the fortunes of the same set of individuals. We are interested, broadly speaking, in how individuals' incomes change over time in a society. "Income mobility" is a shorthand label for this topic. In this chapter, we address questions such as: what exactly do we mean by mobility, and why should we be interested in it? How should mobility be measured? What is the evidence about income mobility for rich industrialized nations?

The period of time over which income mobility is assessed is a fundamental issue, and different choices have led to two relatively distinct literatures. On the one hand, there is the subject of how an individual's income changes between one year and another during their lifetime; on the other hand, there is the subject of income change between generations of parents and children. We use this distinction between intragenerational and intergenerational income mobility as an organizational device in this chapter, reflecting the division in existing literature, but we shall also attempt to draw out the features of the measurement of income mobility that are common to both topics while also highlighting dimensions of them for which different approaches to analysis are appropriate.

Conceptual issues are addressed first because clarification of them is an essential preliminary to any discussion of measurement principles, data sources, and assessment of empirical evidence. In Section 10.2, we review the reasons why and how income mobility is said to be of interest. There are several distinct reasons, and this is because, as we also discuss, there are multiple concepts of mobility, each of which arguably has normative validity. This situation contrasts with assessments of an income distributions at a point in time, in which case there is greater consensus about what is meant by income inequality and how it might be accounted for in social welfare evaluations.

We review the measurement of income mobility in Section 10.3, focusing on the generic case in which there are data on income at two points in time, whether this be two years (as in the intragenerational mobility literature) or two generations (as in the intergenerational mobility literature). This is the most commonly examined situation. Thus we are interested in not only summarizing a single bivariate joint distribution of income but also comparing such distributions across time or countries to say whether mobility is greater or smaller. We explain various descriptive methods for situations in which income data are either continuous or grouped into categories. First we discuss graphical devices and methods that may be used to undertake mobility comparisons without resort to choice of a particular mobility index (so-called dominance checks). Second, we consider scalar indices of mobility ranging from regression coefficients and correlations through to other more specialized developments.

By considering measurement from a generic point of view, we aim to show how there might be greater cross-fertilization between the intra- and intergenerational mobility literatures in approaches to measurement. At the same time, we highlight how the different measurement approaches relate to different concepts of mobility identified in Section 10.2.

Evidence about income mobility is the subject of the next two sections: Section 10.4 considers intragenerational mobility; Section 10.5 considers intergenerational mobility. In each case, our strategy is to build a bridge linking concepts and measurement principles to empirical evidence by first discussing data sources, as well as issues of empirical implementation including data comparability and quality more generally.

The final section, 10.6, provides brief concluding remarks and makes some proposals concerning where the returns to future research efforts are the greatest.

Earlier research on income mobility has typically focused on either within- or between-generation topics. For surveys of intragenerational measurement issues, we build on Jenkins (2011a) who, in turn, draws heavily on other surveys such as by, e.g., Atkinson et al. (1992), Burkhauser and Couch (2009), Fields and Ok (1999a), Jenkins and van Kerm (2009), and Maasoumi (1998). For intergenerational mobility, important earlier reviews are provided by Solon (1999), Björklund and Jäntti (2009), Black and Devereux (2011), and Piketty (2000). Many of the reviews just cited appear in volumes with "Handbook" in their title. Indeed extensive surveys of cross-sectional approaches to income distribution were provided throughout the *Handbook of Income Distribution, Volume 1* (Atkinson and Bourguignon, 2000). It is timely and appropriate to give income mobility similar attention.

Although the chapter draws heavily on the work of others, it also has some distinctive features besides simply being more up-to-date. One aspect is our goal to try to integrate the discussion of intra- and intergenerational mobility insofar as this is possible, while also highlighting what aspects of each topic are intrinsically different and deserving of separate attention. Other aspects include our coverage from conceptual issues through to data, issues of empirical implementation, and evidence.

The emphasis of this chapter is on the *measurement* of income mobility, broadly defined. Of course, it is also of interest to not only describe how individuals' incomes change from one time period to another but also to explain the patterns observed. We have deliberately chosen not to systematically review *models* of mobility to make our task manageable.

There is some discussion of intragenerational models of earnings dynamics, nonetheless, in Section 10.3 because estimates from "variance components" models have been used to derive measures of mobility in the form of income risk. Other types of modeling approaches are reviewed by Jenkins (2000), who also discusses more general issues concerning the modeling of intragenerational income dynamics. These are further elaborated by Jenkins (2011a, chapter 12).

One important distinction is between reduced-form and structural empirical models, each of which has different strengths and weaknesses. The former are empirically grounded rather than derived from a well-developed theoretical model that implies specifications, the parameters of which are estimated from the data. The advantage of a structural approach is that there is a close relationship between parameter estimates and behavioral model parameters, and so interpretation is improved and one may be able to say more about underlying causes. The problem with a structural approach is that clear-cut implications for model specification and proofs of relationships can often only be derived by massive simplification—simplification that compromises claims that the model describes empirical reality. The tension between reduced form and structural approaches has existed for a long time and is likely to remain. The reason for the tension is obvious—approaches combining structure, practicality, and feasibility are very difficult to develop. The problem is that a model is needed not only for the dynamics of labor earnings for an individual but also the earnings and possibly other income sources of other individuals in a multiperson household, and the dynamics of household structure itself also needs to be modeled (Jenkins, 2011a, pp. 368–369).

Exactly the same tension has arisen in empirical modeling of intergenerational income dynamics, where there is also a need to consider not only multiple income sources but also demographic factors. The structural ("optimizing") approach is epitomized by Becker and Tomes (1986) and the reduced-form ("mechanical") approach by a series of papers by Conlisk (1974, 1977, 1984).[1] The relative merits of the two approaches are lucidly discussed by Goldberger (1989), with a "reply to a skeptic" provided by Becker (1989).

10.2. MOBILITY CONCEPTS

Writers on income mobility have long emphasized that mobility has multiple dimensions. For example, a leading survey from a decade ago commented that:

> *the mobility literature does not provide a unified discourse of analysis. This might be because the very notion of income mobility is not well-defined; different studies concentrate on different aspects of this multi-faceted concept. At any rate, it seems safe to say that a considerable degree of confusion confronts a newcomer to the field.*
>
> ***Fields and Ok (1999a, p. 557)***

[1] See also Solon (2004) for a simple model highlighting the key ingredients of an optimizing model and Mulligan (1997) for a monograph-length treatment of the theoretical literature.

The systematic reviews by Fields and Ok and others have done much to reduce the potential confusion. But they cannot banish mobility's multiple facets, and so newcomers continue to require guided tours of the concepts and literature. This section explains what the multiple dimensions of mobility are. We address the question of whether more mobility is socially desirable in each case, arguing that the answer depends on which mobility concept is the focus. A review of the implications of mobility's various facets for social welfare is used to illustrate trade-offs between different types of mobility. We also point out how different concepts have received different emphasis in studies of mobility within or between generations.

10.2.1 Mobility's Multiple Dimensions

Consider first the case in which there are observations on income for N individuals for two periods. In the first period, the income distribution is x, in the second period, the distribution is y; there is a bivariate joint density $f(x, y)$. Overall mobility for the population can be thought of as the transformation linking marginal distribution x with marginal distribution y.

In this section, we distinguish four concepts (Jenkins, 2011a): positional change (which comes in two flavors), individual income growth, reduction of longer-term inequality, and income risk.[2] The different concepts "standardize" the marginal distributions x and y in different ways to focus attention on the nature of the link $x \rightarrow y$.

Positional change refers to mobility that arises separately from any changes in the shapes of the marginal distributions in each period, for example, a rise in average income or in income inequality or, more generally, a change in the concentration of individuals at different points along the income range in y compared to in x. Standardization for such changes is most easily accomplished by summarizing each person's position not in terms of their income per se but in terms of their rank in the population normalized by the population size. (The marginal distribution of these "fractional" (or "normalized") ranks is a standard uniform distribution for both x and y.) Thus positional change mobility refers to the pattern of exchange of individuals between positions, while abstracting from any change in the concentration of people in a particular slot in each year. The latter change is "structural mobility," whereas the former is "exchange mobility": see, for example, Markandya (1984). Changes in income affect positional mobility only insofar as these changes alter each person's position relative to the position of others. Equiproportionate income growth or equal absolute additions to income for everyone raise incomes, but there is immobility in the positional sense.

There are some distinctive characteristics of the concept of mobility as positional change. Mobility for any specific individual necessarily depends on other people's

[2] This classification is similar to that employed by Fields and Ok (1999a) and Fields (2006). See also Van de gaer et al. (2001).

positions as well, which is not true for every mobility concept, as we shall see. The definition of each person's origin and destination position depends on the positions of everyone else in the society: It is these taken altogether that define a hierarchy of positions. Second, and related, if one person changes position, then so too must at least one other person. It is not possible for everyone to be upwardly mobile or, indeed, downwardly mobile. Third, the situation corresponding to "no mobility" is straightforwardly defined: Maximum immobility occurs when every person has the same position in x and in y. If income mobility is summarized using a transition matrix (see Section 10.3) in which cell entries a_{jk} show the probability that an individual in income class j in period 1 is found in income class k in period 2, then maximum immobility is the case in which $a_{jk}=1$ for all income classes (all individuals are on the leading diagonal). However, fourth, there are two different ways of thinking about what reference points to use when there is mobility, one focusing on lack of dependence and the second focusing on movement.

One situation is when one's destination is completely unrelated to one's income origin ("origin independence"). For example, the chances of being found in the richest 10th in period 2 are exactly the same for people who were in the poorest 10th in period 1 as for the people who were in the richest 10th in period 1. In transition matrix terms, this is the case in which $a_{jk}=a_{mk}$ for all origin classes j or m (each row of the transition matrix has identical entries). Another view is that the reference case when there is mobility is if destination positions are a complete reversal of origin positions ("rank reversal"), emphasizing positional movement per se. For example, the poorest person in period 1 is the richest person in period 2, and the richest person in period 1 is the poorest person in period 2, and so on. All entries in the transition matrix lie on the diagonal going from bottom left (richest origin class and poorest destination class) to top right (poorest origin class and richest destination class).[3]

Mobility as *individual income growth* refers to an aggregate measure of the changes in income experienced by each individual within the society between two points in time, where the individual-level changes might be gains or losses. Income growth is defined for each individual separately, and income mobility for society overall is derived by aggregating the mobility experienced by each and every individual.[4] This mobility concept contrasts sharply with the positional change one in several ways. No distinction is made between structural and exchange mobility; it is gross (total) mobility that is described. It is possible for everyone to be upwardly mobile or, indeed, to be downwardly mobile. Positive income growth for everyone may count as mobility even if relative positions

[3] The two reference points are sometimes referred to as cases of "perfect" or "maximum" mobility, but we resist these. The language in the former case makes potentially unwarranted assumptions about the optimality of particular mobility configurations (to be discussed below), and it is difficult to argue that origin independence represents "maximum" mobility in the literal sense.

[4] This is an assumption, albeit commonly made. It is what Fields and Ok (1996) call the "individualistic contribution" axiom.

are preserved. Thus, standardization of the marginal distributions is not an essential feature of the concept.

In the individual economic growth case, it is natural to define mobility for each individual in terms of "distance" between origin and destination income and to think of the maximum immobility case for the population as being when the measure of distance equals zero for every individual ($x_i = y_i$ for all i). Mobility is greater if the distance between origin and destination is greater for any individual, other things being equal. This is similar to the idea of greater movement, meaning more mobility according to the "reversals" version of positional mobility. Again, there is no natural maximum mobility reference point as distance has no obvious upper bound.[5] Defining the metric for "distance" in terms of the income change for each individual is, of course, vitally important for the concept, and the main distinctions have been measures of "directional" and "nondirectional" growth. In the first case, income increases over time are treated differently from income decreases; in the second, an income increase and an income decrease of equal magnitude are attributed the same distance and the measure summarizes income "flux" (more on this shortly). For more precise definitions, see Fields and Ok (1999a).

The third mobility concept defines income mobility with reference to its *impact on inequality in longer-term incomes*. The longer-term income for each individual is defined as the longitudinal average of incomes in each period (variations on this are considered later). In the two-period case, longer-term income equals $^1/_2(x_i + y_i)$ for each i. Averaging across time smooths the longitudinal variability in each person's income, and in addition, the inequality across individuals in these longitudinally averaged incomes will be less than the dispersion across individuals in their incomes for any single period. Mobility can therefore be characterized in terms of the extent to which inequality in longer-term income is less than the inequality in marginal distributions of period-specific income. See Shorrocks (1978a) and later discussion for further details. The zero mobility reference point is when the income of each person in every period is equal to their longer-term income; there is complete rigidity. At the other extreme, maximum mobility occurs when there is inequality in per-period incomes but no inequality at all in longer-term incomes. The issue of whether everyone can be upwardly (or downwardly) mobile does not arise with this mobility concept because it defines mobility using inequality comparisons, and inequality is measured at the aggregate (population) level. There are similarities between this concept of mobility and the rank reversal flavor of the positional change concept because both are concerned with movement, but they use different reference points to assess this (longer-term incomes versus base-period positions, respectively). We return to this issue later.

[5] Individual income growth cannot be represented using a transition matrix because the mobility concept in this case is intrinsically individual rather than group based. However, income growth can be represented using a mobility matrix in which category boundaries are defined in real income terms.

The fourth concept of mobility, as *income risk*, is related to the third. The previous paragraph expressed each person's period-specific income as the sum of a "permanent" component (the longer-term average) and a "transitory" component (the period-specific deviations from the average). Suppose now that the longer-term average is given a behavioral interpretation: It is the expected future income per period given information in the first period about future incomes. From this ex ante perspective, the transitory components represent unexpected idiosyncratic shocks to income, and the greater their dispersion across individuals each period, the greater is income risk for this population. The measure of mobility cited in the previous paragraph (i.e., the inequality reduction associated with longitudinal averaging of incomes) is now reinterpreted as a measure of income risk and has different normative implications (see below). Income movement over time represents unpredictability. This is essentially what Fields and Ok (1999a) refer to as income "flux" (nondirectional income movement). Despite their apparent similarities in construction, the concepts of mobility as inequality-reduction and as income risk diverge in practice when the process describing income generation is not a simple sum of a fixed individual-level permanent component and an idiosyncratic transitory component. Econometric models have been developed with more complicated descriptions of how the permanent and transitory components evolve over time and these imply, in turn, different calculations of expected income and transitory deviations from it. However, the distinction between predictable relatively fixed elements and unpredictable transitory elements of income is maintained and hence so too is a link between mobility as transitory variation and income risk.

10.2.2 Is Income Mobility Socially Desirable?

In what ways are these various mobility concepts of public interest over and above providing useful descriptive content? Does having more mobility represent a social improvement, or is it undesirable? The answers depend on the mobility concept employed, and the support for the different concepts has depended on whether one is assessing within- or between-generation mobility.

Greater mobility in the sense of less association between origins and destinations has long been linked with having a more open society; if where you end up does not depend on where you started from, there is greater equality of opportunity. For example, a classic statement by R. H. Tawney, originally from 1931, is that equality of opportunity

> *obtains in so far as, and only in so far as, each member of a community, whatever his birth, or occupation, or social position, possesses in fact, and not merely in form, equal chances of using to the full his natural endowments of physique, of character, and of intelligence.*
>
> **Tawney (1964, pp. 103–105)**

More recently, a UK government advisor's report on Social Mobility stated that "Social mobility matters because . . . equality of opportunity is an aspiration across the political spectrum. Lack of social mobility implies inequality of opportunity" (Aldridge, 2001, p. 1). For more about equality of opportunity, see Chapter 5.

From this perspective, greater mobility is socially desirable because equality of opportunity is a principle that is widely supported, regardless of attitudes to inequality of outcomes. This is relevant because independence of origins and destinations is consistent with inequality of outcomes being relatively equal or unequal. The argument just rehearsed is, however, typically made in the context of intergenerational mobility rather than intragenerational mobility, and origins refer to parental circumstances, such as "birth, or occupation, or social position" referred to by Tawney. The appeal to fairness in this context is based on the meritocratic idea that someone's life chances should depend on their own abilities and efforts rather than on who their parents were. At the same time, it is important to appreciate that the degree of intergenerational association is an imperfect indicator of the degree of inequality of opportunity.

The degree of origin independence is a direct measure of inequality of opportunity only if two rather special conditions apply (Roemer, 2004). First, the advantages associated with parental background (over which it is assumed that an individual had no choice) are entirely summarized by parental income. Second, the concept of equality of opportunity that is employed views as unacceptable any income differences in the children's generation that are attributable to differences in innate talents (which might be partly genetically inherited). This is what Swift (2006) described as a "radical" interpretation of the equality of opportunity principle and likely to command much less widespread assent than what he refers to as the "minimal" and "conventional" definitions (respectively, access and recruitment processes to life chances are free of prejudice and discrimination; and outcomes achieved depend on "ability" and "effort" but not on family background).

The social desirability of mobility as independence of origins has less force in the intragenerational context. The reason is that incomes are measured at a point within the life course. By that stage, period-1 incomes are likely to reflect differences in peoples' abilities and efforts (in addition to family background and other factors), and period-2 incomes to reflect the persisting effects of these factors. To the extent that abilities and efforts do play this role (or are seen to) and also viewed as fair on the grounds of merit or desert, the reduction of dependence between origins and destination has less appeal as a principle of social justice.

More common in the within-generation context are statements that income mobility is desirable because it is a force for reduction in the inequality of longer-term incomes. The most famous statement in this connection was by Milton Friedman six decades ago in his *Capitalism and Freedom* (though observe that he also refers to equality of opportunity in this context):

> *A major problem in interpreting evidence on the distribution of income is the need to distinguish two basically different kinds of inequality; temporary, short-run differences in income, and*

> *differences in long-run income status. Consider two societies that have the same annual distribution of income. In one there is great mobility and change so that the position of particular families in the income hierarchy varies widely from year to year. In the other there is great rigidity so that each family stays in the same position year after year. The one kind of inequality is a sign of dynamic change, social mobility, equality of opportunity; the other, of a status society.*
>
> **Friedman (1962, p. 171)**

Similar views are apparent across the political spectrum in the United States. The chairman of President Obama's Council of Economic Advisors recently stated,

> *Higher income inequality would be less of a concern if low-income earners became high-income earners at some point in their career, or if children of low-income parents had a good chance of climbing up the income scales when they grow up. In other words, if we had a high degree of income mobility we would be less concerned about the degree of inequality in any given year.*
>
> **Krueger (2012, p. 3)**

Although both authors are referring to the distributions of incomes within generations, one could extend the same inequality-reduction idea to the intergenerational context, by summarizing mobility in terms of the extent to which dynastic inequality (referring to incomes averaged over generations of the same family) is less than the inequality in any given generation. But this is rarely done, perhaps because the normative appeal of the dynastic average income is much less than that of a multiperiod average within generations, and data for more than two generations are rarely available.

According to the arguments about longer-term inequality reduction, income mobility is socially desirable for instrumental reasons rather than for its own sake. That is, society is assumed to care about income inequality (less is better, other things being equal), but inequality is assessed using longer-term incomes, and year-to-year mobility means that the inequality of this distribution is less than the inequality of incomes in any particular year. The normative content of the mobility principle therefore hinges on views concerning the nature and validity of the benchmark that is provided by the distribution of longer-term incomes. As Shorrocks points out,[6] there is

> *the presumption that individuals are indifferent between two income streams offering the same real present value. This might be true if capital markets were perfect (or if there was perfect substitutability of income between periods), but it seems likely that individuals are concerned with both the average rate of income receipts and the pattern of receipts over time. We may go further and suggest that individuals tend to prefer a constant income stream, or one which is growing steadily, to one which continually fluctuates.*
>
> **Shorrocks (1978a, p. 392)**

[6] Shorrocks also draws attention to the assumption that the same measure is used to summarize both the dispersion of longer-period incomes and the dispersion of per-period incomes.

Thus, the argument is not only about the feasibility of smoothing incomes to achieve the longer-term average, but also the undesirability of the uncertainty associated with a fluctuating income stream.

This brings us to the fourth concept of income mobility, as income risk. To illustrate this, Shorrocks defines for each individual a "constant income flow rate generating receipts which gives the same level of welfare as the income stream he currently faces" (Shorrocks, 1978a, p. 392), and he argues that

> *[r]eplacing actual recorded incomes with this alternative income concept in the computation of inequality values introduces a new dimension into the discussion of mobility. No longer is mobility necessarily desirable. Changes in relative incomes still tend over time to equalise the distribution of total income receipts, and to this extent welfare is improved. But greater variability of incomes about the same average level is disliked by individuals who prefer a stable flow. So to the extent that mobility leads to more pronounced fluctuations and more uncertainty, it is not regarded as socially desirable. A more detailed examination of these two facets of mobility will provide a better understanding of the impact of income variability and the implications for social welfare.*
>
> **Shorrocks (1978a, pp. 392–393)**

Thus, even though income mobility has an inequality-reducing impact, mobility is not necessarily socially desirable if mobility represents transitory shocks. In this case, mobility is a synonym for not only income fluctuation but also unpredictability and economic insecurity. Fluctuating incomes are undesirable because most people prefer greater stability in income flows to less, other things being equal, if only because it facilitates easier and better planning for the future. But, more than this, by definition, transitory income variation is an idiosyncratic shock that cannot be predicted at the individual level; greater transitory variation corresponds to greater income risk, and greater risk is undesirable for risk-averse individuals. The definition of the "alternative income concept" from which transitory shocks deviate is, of course, crucial, and we return to this.

What about the social desirability of individual income growth (the second mobility concept)? The answer is not clear cut because it depends on the nature of the income growth and who receives it. An increase in income for any given individual is a social improvement, and an income fall is socially undesirable. The main issue, then, is how to aggregate gains and losses in the social calculus. Evaluation of the impact of individual income growth on the welfare of society as a whole requires a weighing up of the gains and losses for different people, and opinions are likely to differ about how to do this. An egalitarian may weight income gains for the initially poor greater than income gains for the initially rich because this will contribute to reducing income differences between them over time. (On the progressivity of income growth, see, e.g., Benabou and Ok, 2000 and Jenkins and van Kerm, 2006.)

Arguments to the contrary appealing to principles of desert or incentives might also be made. It might be argued, for instance, that differential income growth rates are of less concern if income gains among the rich reflect appropriate returns to entrepreneurial activity or to widely acclaimed talents. The rise in bankers' bonuses in the manner observed in many Anglophone countries in recent years may not count as an example of the former. But as an example of the latter, we note the views of the UK's former Prime Minister Tony Blair expressed in an interview asking him whether it was acceptable for the gap between rich and poor to get bigger. His response referred instead to individual income growth:

> *[T]he justice for me is concentrated on lifting incomes of those that don't have a decent income. It's not a burning ambition for me to make sure that David Beckham earns less money, . . . [T]he issue isn't in fact whether the very richest person ends up becoming richer . . . the most important thing is to level up, not level down.*
>
> ***Interview on BBC Newsnight (5 June 2001)***[7]

Another concept of desert may also be relevant when assessing mobility. This is the argument concerning "distressed gentlefolk"—people who were previously well-off, but experience a significant fall in resources through no fault of their own. Thus income gains and income losses for an individual may not be assessed symmetrically but, again, relate to why income changed (see also the discussion of "loss aversion" below).

We end this section with two observations. First, our discussion of the social desirability or otherwise of income mobility has referred to income movement from throughout the range of base-period income origins to all potential final-period income destinations. There has been no particular focus on persistence at the bottom or at the top. In part, this is because such a focus arguably does not raise additional conceptual issues, except where to draw the cutoffs demarcating the poor and nonpoor, or rich and nonrich. Indeed if the bivariate joint distribution is summarized using a transition matrix, then suitable definition of the income groups reveals the movement at the top and the bottom. However, we do discuss selected aspects of the measurement of high- and low-income persistence in the next two sections.

Second, our discussion of the social desirability of mobility has focused on its normative aspects. We ignore the positive political economy arguments about public support for mobility. On this, see, e.g., the analysis by Benabou and Ok (2001) of the "prospect of upward mobility" (POUM) hypothesis, which is that individuals who currently have low income may not support high levels of redistribution because of their aspiration that they or their children will become rich in the future.

10.2.3 Income Mobility and Social Welfare

The discussion so far demonstrates that the impact on social welfare of greater income mobility is not clear cut and depends on the mobility concept that is emphasized. A natural question for an economist to ask is whether there are explicit welfare

[7] Transcript at http://news.bbc.co.uk/1/hi/events/newsnight/1372220.stm.

foundations for the various mobility concepts that have been discussed so far. For inequality measurement, the use of an explicit model of social welfare is known to yield dividends; see, notably, Atkinson's (1970) demonstration of how the "cost" of income inequality can be summarized in social welfare terms and how inequality comparisons based on Lorenz curves are intimately linked to orderings by social welfare functions (SWFs) that are additive, increasing, and concave functions of individuals' incomes. The corresponding literature on the social welfare foundations of mobility measurement is small, with contributions including Atkinson (1981a), reprinted as Atkinson (1983), Atkinson and Bourguignon (1982), Markandya (1984), and Gottschalk and Spolaore (2002). In this section, we focus on the nature of the SWFs employed in the mobility context; how these functions relate to mobility dominance results is discussed later.

The SWF used in the multiperiod context is a straightforward generalization of the one-period case discussed by Atkinson (1970). Overall social welfare, W, is the expected value (average) of the utility-of-income functions of individuals. In the two-period case, the utility-of-income function is $U(x,y)$ and weighted by the joint probability density $f(x,y)$. That is,

$$W = \int_0^{a_y}\int_0^{a_x} U(x, y) f(x, y)\mathrm{d}x\mathrm{d}y, \tag{10.1}$$

where $U(x,y)$ is differentiable and a_x and a_y are the maximum incomes in periods 1 and 2. It is assumed that increases in income in either period are desirable, other things being equal (so positive income growth raises utility): $U_1 \geq 0$ and $U_2 \geq 0$.

Research in this tradition concentrates on the case in which the marginal distributions x and y are identical. In other words, the economic context is the same as the one used earlier to characterize positional mobility. All relevant mobility is encapsulated by the changes in individuals' ranks or by the transition matrix when individual incomes are classified into discrete classes. Atkinson and Bourguignon (1982) show that if the SWF is additively separable across time periods (so that $U_{12} = 0$, then income mobility is irrelevant for social welfare; only the marginal distributions matter.[8] If, instead, $U(x,y)$ is a concave transformation of the sum of the per-period utilities, then $U_{12} < 0$.

How does one interpret this sign? Atkinson and Bourguignon (1982) discussed the class of least concave functions associated with a particular preference ordering and the special case in which preferences are homothetic. In this situation, the utility function $U(.)$ is neatly characterized by two parameters: $\varepsilon > 0$ summarizing *aversion to inequality of multiperiod utility*, and $\rho > 0$ summarizing the inverse of the elasticity of substitution between income in each period (i.e., the degree of *aversion to intertemporal fluctuations in income*; Gottschalk and Spolaore, 2002, p. 195). The case $U_{12} < 0$ corresponds to

[8] See also Markandya (1984) and Kanbur and Stiglitz (1986).

the situation in which $\varepsilon > \rho$, i.e., in the social welfare assessment, multiperiod inequality aversion offsets aversion to intertemporal fluctuations (which are of course reducing multiperiod inequality). When $\rho = 0$, an increase in income mobility must increase social welfare. With perfect substitution of income between periods, one is only interested in the reduction of multiperiod inequality.

Gottschalk and Spolaore (2002) pointed out that origin dependence has no role in the Atkinson–Bourguignon model.[9] In transition matrix terms, if there is any preference at all for income reversals ($\varepsilon > \rho$), not only does an increase in mobility represent a social welfare gain, but also the complete reversal scenario is preferred to the origin independence one. This feature has relevance to the application of the social welfare framework to mobility measurement using stochastic dominance checks (discussed in the next section). The irrelevance of origin dependence suggests that the approach is less applicable to intergenerational mobility comparisons because origin independence is the principle most commonly espoused in that context (see earlier discussion).

However, an important contribution of Gottschalk and Spolaore (2002) was to show that greater origin independence can be social welfare improving if the SWF is generalized to take account of aversion to future income risk. In the two-period context, they drop Atkinson and Bourguignon's assumption that period-2 income is known with certainty in period 1. Individuals take conditional expectations of period-2 incomes based on observed period-1 incomes and the joint density of outcomes. With homothetic preferences, the utility function is now characterized by a third parameter, γ, summarizing the degree of aversion to second-period risk. As Gottschalk and Spolaore demonstrated,

> *Origin independence reduces both multi-period inequality and intertemporal fluctuations, but increases future risk. Individuals will positively value origin independence as long as aversion to multi-period inequality and aversion to fluctuations dominate aversion to future risk (ε and ρ are not smaller than γ, and at least one of them is larger).*
>
> ***Gottschalk and Spolaore (2002, p. 204)***

In summary, evaluation of income mobility in terms of social welfare has payoffs. There is a single unifying framework. Within this, whether an increase in income mobility is social welfare improving depends on the priority given to different mobility concepts. For instance, reversals are less likely to be valued the greater the aversion to intertemporal fluctuations and to future income risk, but more likely to be valued the greater the aversion to multiperiod inequality. Nonetheless one limitation of the SWF framework discussed so far is that it does not incorporate evaluations of mobility in the form of individual income growth—apart from aspects of this that overlap with the other concepts. One leading exception is the research by Bourguignon (2011), who shows that the Atkinson and Bourguignon results can be applied to comparisons of alternative "growth

[9] See also similar remarks by Fields and Ok (1999a, pp. 578–579).

processes" in the case in which the pair of marginal distributions relating to the first period are identical. However, this is a severe constraint on the applicability of the results.

An alternative strategy is to define SWFs explicitly in terms of income mobility—income changes rather than income levels. For example, one may assume that individual-level mobilities are represented by some measure of "distance" between first- and second-period incomes for each individual i, $d(x_i,y_i)$, where the distance function is common to all individuals, and a social weight. Overall social welfare is the weighted sum over individuals of the d_i. King (1983) and Chakravarty (1984) assume that d_i is a function of period-1 and period-2 income ranks (the positional mobility case), and that reranking is desirable ($\partial W/\partial d_i > 0$) and the social weight is increasing in period-2 income. By contrast, for Van Kerm (2006, 2009) and Jenkins and van Kerm (2011), d_i is a directional measure of individual income growth, and the social weight depends on base-year income ranks. For a more general discussion, see Bourguignon (2011), who discussed how the Atkinson–Bourguignon utility-of-income function, $U(x,y)$, can be rewritten as $V(x,y-x)$ with the same properties on the differentials of the second (income change) argument. This framework would lead one to question, for example, the approach of Fields et al. (2002), whose SWF is the simple average of the d_i (equality of social weights), and so $\partial V/\partial x = 0$: mobility evaluations do not depend on initial income at all.

The main advantage of defining SWFs in terms of mobility directly is that there is great flexibility in the specification of the distance function d_i. The disadvantage of the approach is that it runs the risk of being ad hoc rather than a general unifying framework like the Atkinson and Bourguignon (1982) one. In particular, how should the social weights be specified? Unfortunately, the Bourguignon (2011) framework provides no simple answers.

The social welfare approaches described so far assume that W is a form of expected utility evaluation, though modified to context: Atkinson and Bourguignon (1982) incorporated preferences that were not time-additive, and in addition, Gottschalk and Spolaore (2002) abandoned complete predictability of income. A different approach altogether is to suppose that evaluations are based not on expected utility but prospect theory. Jäntti et al. (2014) explored this idea, utilizing a utility function that incorporates reference-income dependence and loss aversion. The latter feature means that, over and above any preference for smooth rather than fluctuating incomes over time, fluctuations lower individuals' welfare directly because losses outweigh gains of equal size. There is therefore an asymmetric treatment of income decreases and decreases, as for the "distressed gentlefolk" argument cited earlier but rather differently motivated. This approach is a promising area of research and chimes with more popular expressions of the problem of growing income risk. Hacker and Jacobs (2008), for instance, specifically cited loss aversion as one of the factors related to the growth of income risk in the United States.

10.3. MOBILITY MEASUREMENT

This section is about measuring mobility. First we discuss descriptive devices, by which we mean graphical and tabular methods for summarizing patterns of mobility. We consider them in more detail than other surveys because we think it is important to "let the data speak" (though there are limits to which this is possible, as we show). Second, we describe how descriptive devices also have normative implications, being linked to dominance checks for mobility comparisons. Third, we consider scalar indices of mobility. Throughout the section we relate the descriptive devices and measures to the different concepts of mobility identified earlier. Most of the examples that we use are drawn from the intragenerational literature, reflecting their greater use in that context. But one of the lessons to be drawn is that the same methods could also be applied to the intergenerational context.

10.3.1 Describing Mobility

In the two-period case, the bivariate joint distribution of income contains all the information there is about mobility, so a natural way to begin is by summarizing the joint distribution in tabular or graphical form.[10] How one proceeds depends on the nature of the data to hand and the mobility concept of interest. We have been assuming that income distributions are continuous but in practice it is often convenient to represent the data in grouped form, or the data may be intrinsically discrete as in the case of "social classes." In addition the information content of the descriptive device is related to the way (if any) in which the analyst standardizes the marginal distributions of any one bivariate distribution and, when making comparisons of bivariate distributions, makes further adjustments (e.g., to control for differences in average income between the bivariate distributions for two countries). If one is solely interested in pure exchange mobility (changes in relative position), then both issues are dealt with by working with the fractional rank implied by an individual's income rather than the income itself. In this case, all the marginal distributions are standard uniform variates and the same across time periods and countries.[11] But if the focus is on other mobility concepts, other standardizations may be used.

A mobility matrix, M, is constructed by first dividing the income range of each marginal distribution into a number of categories (which need not be the same in each period, but typically is) and cross-tabulating the relative frequencies of observations with each

[10] We consider a summary device for mobility as equalization of longer-term income in the case when there are more two periods in the next subsection.

[11] Fractional (or "normalized") ranks range between zero and one, with a mean of 0.5. Particular care needs to taken in their estimation when there are tied income values to ensure that these conditions are met. See, for example, Lerman and Yitzhaki (1989).

matrix cell: typical element $m_{i\,j}$ is the relative frequency of observations with period 1 income in range (group) i and period-2 income in range j. The graphical representation of the discrete joint probability density function is the bivariate histogram. Alternatively, the mobility process may be represented by the transition matrix and the marginal distributions. Borrowing notation from Atkinson (1981a), suppose that there are n income ranges, with the relative number of observations in group k in period-1 is m_1^k for $k=1, \ldots, n$, and correspondingly in period 2. The marginal (discrete) distribution in period-1 is summarized by the vector $m_1=(m_1^1, m_1^2, \ldots, m_1^n)$ and correspondingly for period-2. Hence,

$$m_1^k = m_2^k A. \tag{10.2}$$

When the focus is on pure exchange mobility, the ranges typically refer to quantile groups. For example, in the case of decile groups, each group contains one-tenth of the population. The transition matrix is then bistochastic. Mobility is entirely characterized by the transition matrix A.

An illustrative example is shown in Table 10.1. Mobility refers to changes in the relative positions in the United States between 1979 and 1988, and 1989 and 1998, with each individual's income defined as the equivalized real annual family disposable income of the family to which the individual belongs. The United States in the 1980s and the 1990s is a long way from the total immobility scenario (in which every cell percentage would equal to zero, except those on the leading diagonal, which would equal 100%). Clearly, there is also neither origin independence (every cell entry equal to 10%) nor total reversal of positions. The general pattern is one of much short-distance mobility with long-distance mobility being rare. For example, of those individuals in the poorest 10th in 1989, around 42% are also in the poorest 10th in 1998 with fewer than 1% making it to the richest 10th. Of the richest 10th in 1989, around 46% stay in that group, and less than 2% are in the poorest 10th in 1998. More generally, the largest transition proportions are on or close to the matrix diagonal (Hungerford (2011) reported that 73% of individuals remained in the same 10th or moved at most two deciles), and upward and downward mobility appears to be broadly symmetric. Because the U.S. situation described in Table 10.1 is not particularly close to the standard mobility reference points, it is not straightforward to say whether there is a large or small amount of mobility. It is also of interest to assess whether mobility increased between the 1980s and 1990s. Methods for mobility comparisons are discussed in the measurement section that follows. Further empirical evidence about within-generation mobility is presented in Section 10.4.

If the interest is in mobility other than of the positional kind, changes in the marginal distributions are also of interest. A particular example might be when the income class boundaries are defined as fractions of median income, or as fractions of the poverty line and there is interest in poverty rate trends as well as movements into and out of low

Table 10.1 Decile transition matrices: United States, (a) 1979–1988 and (b) 1989–1998 (percentages)

Origin group	Destination group 1	2	3	4	5	6	7	8	9	10
1979					**1988**					
1	44.3	18.3	12.4	9.2	7.1	3.0	1.8	2.0	0.7	1.3
2	18.1	25.3	21.0	11.7	7.5	5.4	4.7	3.2	1.9	1.1
3	10.6	18.2	15.3	16.8	11.6	9.0	8.8	4.9	3.1	1.7
4	7.2	8.9	14.0	14.0	14.7	15.7	12.0	5.6	6.0	2.1
5	6.1	9.2	10.9	12.8	13.3	16.9	12.3	7.5	7.7	3.4
6	4.1	5.2	8.8	10.3	11.8	10.0	14.2	16.9	12.6	6.2
7	3.5	6.5	6.9	8.6	10.4	13.4	13.3	16.8	13.4	7.2
8	3.1	4.6	3.2	7.7	12.3	9.5	12.6	15.7	17.7	13.6
9	1.2	2.2	4.8	6.3	6.9	10.2	12.2	14.7	18.0	23.5
10	2.1	1.5	2.8	2.5	4.2	7.0	8.5	12.8	18.6	40.0
1989					**1998**					
1	41.9	21.6	13.7	7.0	4.6	3.7	2.7	2.2	1.9	0.7
2	20.4	22.5	15.4	11.6	11.0	8.1	4.0	4.0	1.7	1.2
3	12.5	20.8	17.1	16.4	10.9	10.3	5.2	3.2	1.7	1.9
4	6.9	11.6	15.5	16.9	14.5	11.4	10.1	7.7	2.3	3.1
5	4.8	6.2	12.2	13.8	16.0	14.2	12.4	7.1	7.5	5.8
6	3.2	3.7	9.1	11.6	16.0	14.4	15.7	11.7	7.7	6.9
7	3.2	4.5	7.6	9.3	8.7	12.2	16.3	15.6	16.8	5.8
8	3.0	4.7	5.2	5.4	7.9	12.1	17.2	17.0	19.3	8.3
9	2.5	3.1	4.0	4.9	7.5	7.1	10.7	18.2	21.8	20.3
10	1.7	1.0	0.4	3.2	3.0	6.3	6.0	13.1	19.3	46.1

Note: Income refers to equivalized real annual family disposable income, distributed among all individuals (adults and children). The decile groups are ordered from poorest (1) to richest (10).
Source: Hungerford (2011, Tables 2 and 3), based on PSID data.

income.[12] More generally, defining income group boundaries that are fixed in real income terms over time provides indications about individual income growth for individuals of different origins; if each period's incomes are standardized by period-average income, the information refers to income growth relative to the average.[13] (We say "indications" regarding this mobility concept because its essence refers to income changes at the individual rather than group level.) Similarly, the dispersion across origin groups of individuals from a common income origin may be indicative of income risk, but the connection is not altogether obvious. Neither mobility matrices of this kind nor

[12] For examples, see, e.g., Hungerford (1993, 2011) for the United States and Jarvis and Jenkins (1998) for the UK.
[13] For examples, see Hungerford (1993), Hungerford (2011), and Jarvis and Jenkins (1998).

conventional transition matrices are directly informative about mobility as longer-term inequality reduction.

Graphical summaries can complement and sometimes be more effective than tabular presentations; visual impact matters. Even transition matrices and comparisons of them can be visualised. We refer, for instance, to the use of transition probability color plots introduced by Van Kerm (2011). Suppose individuals are classified into vingtile groups in each of period 1 and period 2. For the visualization, individuals are classified according to their income group in period 2, and lined up in rows with the poorest twentieth in one row at the top, the next twentieth in the row beneath, and so on down to the final row containing the richest twentieth. Each person is also tagged with their period-1 group membership using a color coding system. Suppose the poorest twentieth in period 1 is represented by blue and the richest twentieth by red, and the intermediate groups are represented by the colors of the rainbow in between. If there were no changes in relative position over time, every one would remain in their period-1 income group; there would be a one-to-one correspondence between rows and colors. (Rows would consist of full blocks of the same color.) If there were no association between income origin and income destination, every color would form an equal-sized block in each and every row. If there were complete rank reversal, the original color scheme would be reversed, with the richest period-1 group (red) in the top row and the poorest period-1 group (blue) in the bottom row.

Examples of such representations, due to Van Kerm (2011), are shown in Figure 10.1 for individuals' household income mobility between 1987 and 1995 in Western Germany (left) and the United States (right). It is immediately apparent that, over this 12-year period, there is substantial income mobility in both countries and throughout the income distribution, including a small fraction of the richest twentieth falling to

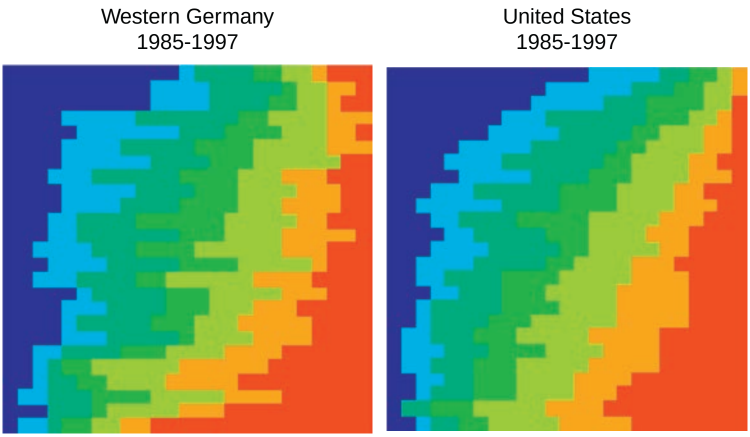


Figure 10.1 Transition color plot examples. *Source: Van Kerm (2011)*. (For color version of this figure, the reader is referred to the online version of this book).

the poorest twentieth, and vice versa. But there is clearly no origin independence in either country, let alone complete rank reversal. Interestingly, however, it is also clear that the main differences in patterns of mobility are at the bottom of the income distribution (more changes in relative position in Western Germany than in the United States). We return to this finding in the next section. The particular advantage of the transition color plots is their visual immediacy. However, color is not always available. The color transition plots summarizing income mobility in the book by Jenkins (2011a, Figure 5.1) were reproduced in black and white, and this reduced their effectiveness.

What about alternative devices? Perhaps the most straightforward way to summarize a bivariate joint distribution is using a scatter plot of period-2 incomes against period-1 incomes. Figure 10.2 provides a within-generation example using British income data for 1991 and 1992.

The advantages of the scatter plot are that it is very easy to produce and provides an immediate impression about the degree of immobility of incomes (the clustering around the 45° line), as well as the nature of the marginal distributions. For a focus on changes in relative position alone, the corresponding scatter plot would be of individuals'

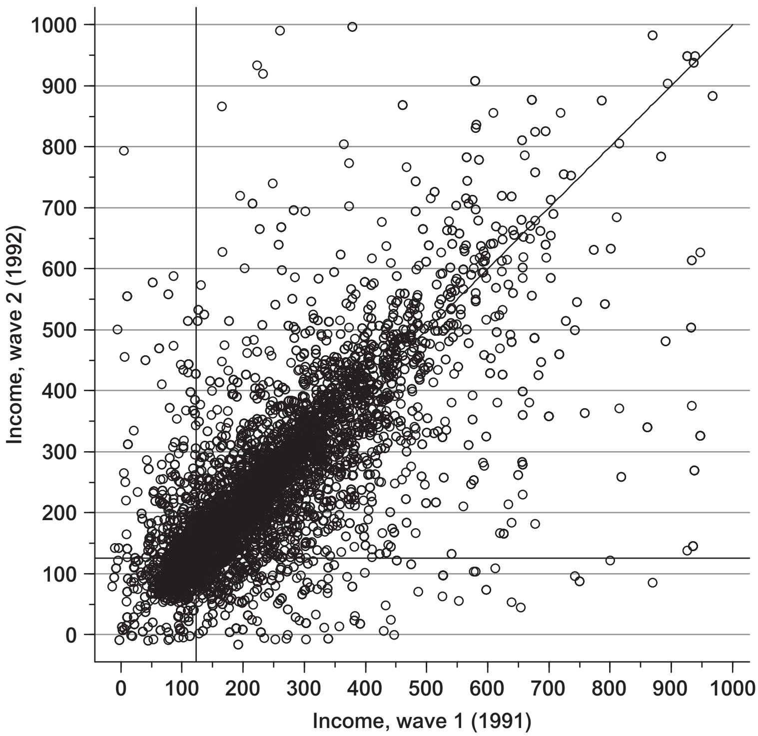


Figure 10.2 Scatter plot example. *Source: Jenkins (2011a, Figure 1.2).*

normalized ranks in each of the two periods. The main disadvantage is that potentially important detail is lost because the bivariate density is not estimated; there is no difference to the eye between 10 observations with a particular combination of period-1 and period-2 incomes and 100 observations with the same pair of incomes.

One way to proceed is derive and plot the joint density. The simplest estimates to produce are those of the bivariate discrete density (essentially plotting the bivariate histogram—see above). However, there are well-known disadvantages of such discretization: As in the univariate distribution case, the estimates are sensitive to choice of income class boundaries, and of course, information within the ranges is lost with the grouping. Kernel density estimation methods avoid the problem because of the way in which they smooth data within a moving window rather than within fixed categories. Figure 10.3 shows a "typical" joint bivariate density for West German family incomes for 2 consecutive years over the period 1983–1989.[14] Incomes in each year are normalized by the contemporaneous median, but otherwise, the marginal distributions are not constrained to be the same (so this is a representation of exchange mobility alone). Compared to the scatter plot, the concentration of individuals on and around the 45° representing perfect immobility is readily apparent. However, the fine detail remains difficult to

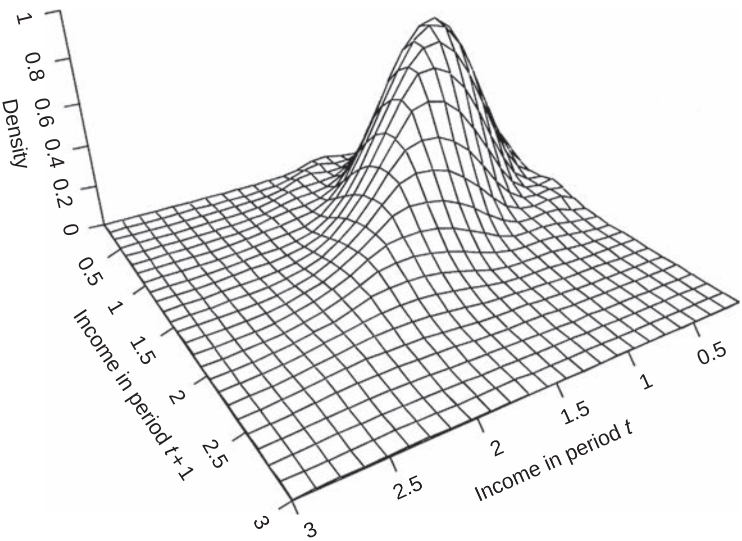


Figure 10.3 Bivariate density plot example. *Note: The charts shows a "typical" kernel density estimate for incomes in two consecutive periods. Source: Schluter (1998, Figure 1).*

[14] The source does not state which specific pair of years over the period 1983–1989 was used for the calculations.

ascertain, partly because the three-dimensional representation has to use a specific projection. What a reader perceives may change if the estimates are viewed from a different angle. Related, differences in marginal distributions are difficult to examine; so too is individual income growth. A further issue, shared with the scatter plot and bivariate histogram, is that it is difficult to compare a pair of bivariate distributions (e.g., for two different countries), even if the plots to be compared are placed adjacent to other. Overlaying one plot on another is far too messy, but without some form of overlay, detailed comparisons are constrained.

Both issues are resolved to some extent by summarizing the density estimates using contour plots in which contour lines connect income pairs with the same density. An example is provided using U.S. and West German income data for 1984 and 1993 in Figure 10.4. Income refers to the log of equivalized family income expressed as a deviation from the national contemporaneous mean. Contour lines are drawn at values that separate the quintile groups for each country (the 20th, 40th, 60th, and 80th percentiles). The solid lines are for the United States, the dotted lines are for West Germany (WG).

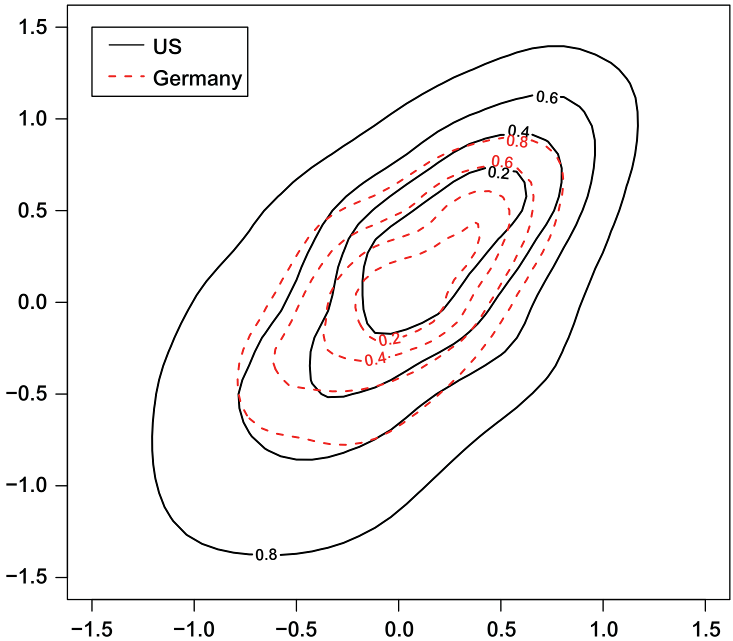
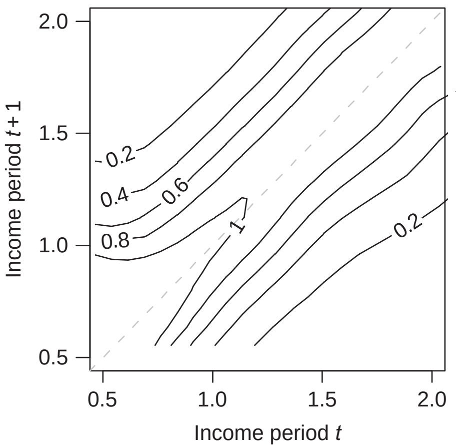


Figure 10.4 Contour plot example. *Note: The chart shows the kernel-smoothed joint density of income in 1984 and 1993 for the United States and West Germany, where income is posttax posttransfer family income equivalized by the PSID equivalence scale, and income for each year is expressed as a deviation from the year-specific mean. Source: Gottschalk and Spolaore (2002, Figure 1), redrawn by the authors.*

As Gottschalk and Spolaore (2002) commented, the plot reveals multiple features of the joint distribution. Each contour line for Germany lies inside its U.S. counterpart, indicating greater cross-sectional inequality in the United States. Clustering around the 45° immobility line is apparent for both countries but is greater for the United States. Also, the contour lines are generally flatter for Germany, meaning that expected period-2 income (conditional on period-1 income) varies less with period 1 in WG than it does in the United States. Gottschalk and Spolaore (2002) commented that this suggests a lower cross-period correlation in the United States, and they also pointed to a greater variation around the conditional means in the United States. Contour plots are also used in the U.S.–West German comparisons by Schluter and Van de gaer (2011, Figure 2).

Just as contour plots for continuous income distributions correspond to mobility matrices, there are also devices for continuous incomes corresponding to the transition matrix. One requires estimates of the conditional density $f(y|x)$, which is straightforwardly estimated in principle using the fact that $f(y|x) = f(y,x)/f(x)$. Estimates of the numerator and denominator are derived across a grid of values of x and y using kernel density estimation. See Quah (1996), who refers to this concept as a "stochastic kernel" and applications to income mobility include Schluter and Van de gaer (2011). Compared to unconditional joint density plots, the conditional density plots allow a more direct comparison of expected income growth across the base year income range. Examples are provided in Figure 10.5 based on data for the United States (top chart) and Western Germany (bottom chart) for 1987 and 1988. Income is equivalized net household income expressed relative to the 1987 median. Schluter and Van de gaer (2011, p. 11) pointed to not only the greater spread of contours in the United States indicating differences in marginal distributions, but also that the "particular … feature of the conditional densities is the greater upward mobility of low-income Germans" compared to low-income Americans. Note the more distinct upturn of the contours in the top left of the Western German chart compared to the shape of the corresponding U.S. contours.

Conditional densities are not the same as conditional probabilities, which is what constitute the transition matrix. Estimation of the conditional (cumulative) probability density $F(y|x)$ requires integration over the marginal distribution of y. As Trede (1998) explained, estimates of $F(y|x)$ can be inverted to give the probabilities for second-period income conditional on particular values of first-period income ("p-quantiles"). Trede's device for "making mobility visible" is a plot of these p-quantiles against first-period income values. Figure 10.6 shows one of these nonparametric transition probability plots using data for West German equivalized family incomes in 1984 and 1985. Incomes are normalized by the 1984 median, so "growth mobility is not excluded from the analysis" (Trede, 1998, p. 80). In the extreme case of origin independence, each transition probability contour would be horizontal. If, instead, there were complete immobility so that second period incomes were completely determined by first period incomes, the contours would lie on top of each other. (In particular, if there were no change in median income,



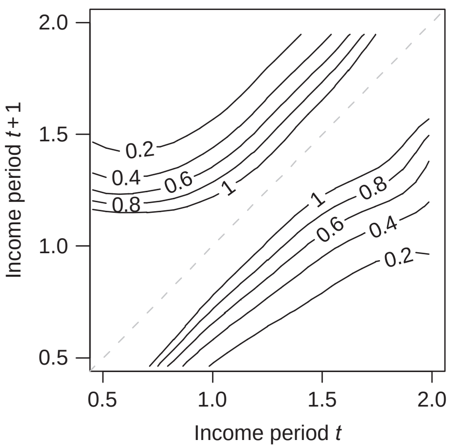


Figure 10.5 Conditional density plot example. *Note: Year t refers to 1987; year t+1 refers to 1988. The top chart refers to the United States; the bottom chart to Western Germany. Source: Schluter and Van de gaer (2011, Figure 2).*

the contours would lie on the 45° line.) The greater the gaps between the contour lines, the greater is inequality in the second period. The slope of the contours is generally less than 45°, indicating some regression to the median. Figure 10.6 shows that, among individuals with median income in 1984, around 10% have an income less than 0.7, and about 10% have an income of at least 1.7 of the 1984 median in 1985. Methods closely related to Trede's are used by Buchinsky and Hunt (1999) to derive nonparametric estimates of transition probability estimates, which the authors reported in tabular rather than chart form.

Patterns of mobility in the form of individual income growth are not shown directly in the devices discussed so far. The simplest way to focus on this aspect is to define income

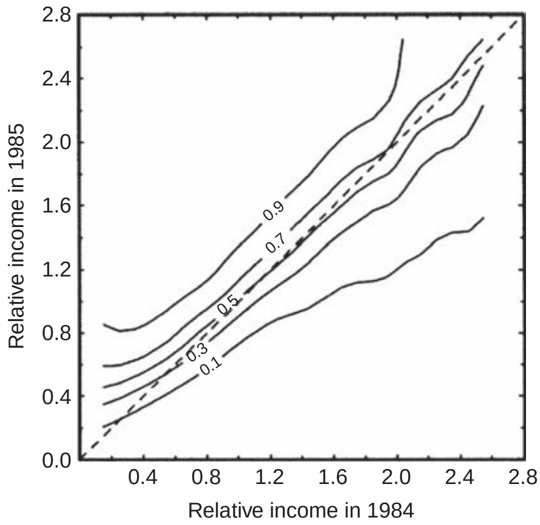


Figure 10.6 Nonparametric transition probability plot example. *Note: Relative income in each year equal to income divided by the 1984 median income. Source: Trede (1998, Figure 1).*

growth at the individual level between the two periods using some measure of directional income growth (Fields and Ok, 1999b), thereby converting the bivariate joint distribution to a univariate distribution of income changes. Then all the devices commonly used for summarizing univariate income distributions are available with one important proviso. Income changes may be negative or zero and not restricted to positive values (and the mean change may also be zero or negative). However, the ratio of second-period income to first-period income is positive (assuming incomes are positive), and it is often convenient to use this metric. Schluter and Van de gaer (2011, Figure 2) present kernel density estimates of the distribution of income ratios. Comparisons based on plots of cumulative distribution functions (CDFs) of income change distributions are also presented by Chen (2009, Figure 4) and Demuynck and Van de gaer (2012, Figure 1).

A CDF plot of this type is based on an ordering of individuals' income *changes* from smallest (most negative) to the largest. One is often interested in the extent to which individual income growth is "pro-poor," that is, whether income growth is greater for those at the bottom of the first-period income distribution relative to those at the top. In particular, pro-poor growth between two periods is a factor reducing the the inequality of second period incomes relative to first period incomes.[15] See also the discussion of SWFs in Section 10.2. Fields et al. (2003) plotted the average change in log

[15] But pro-poor growth does not guarantee inequality reduction. It also leads to reranking, which may have an offsetting effect. See Jenkins and Van Kerm (2006) for a fuller explanation and empirical examples.

per capita income between two time points against income in the base year, for four countries. Comparisons across countries are constrained by the fact that the income range on the horizontal axis (base-year income) varies tremendously. Comparability is enhanced if, instead, one plots individuals' average income change against their normalized (fractional) rank in the base-year distribution (with individuals ordered from poorest to richest). The horizontal axes in this case are bounded by 0 and 1. Such plots were developed by Van Kerm (2006, 2009) and independently by Grimm (2007). Extensive empirical examples are provided by Jenkins and van Kerm (2011) for four 5-year periods in Britain during the 1990s and 2000s, from which Figure 10.7 is taken. (Individual income growth refers to the change in the log of individuals' household income between 2 years.) It is clear that income growth is distinctly pro-poor in each of the subperiods, especially 1998–2002.[16]

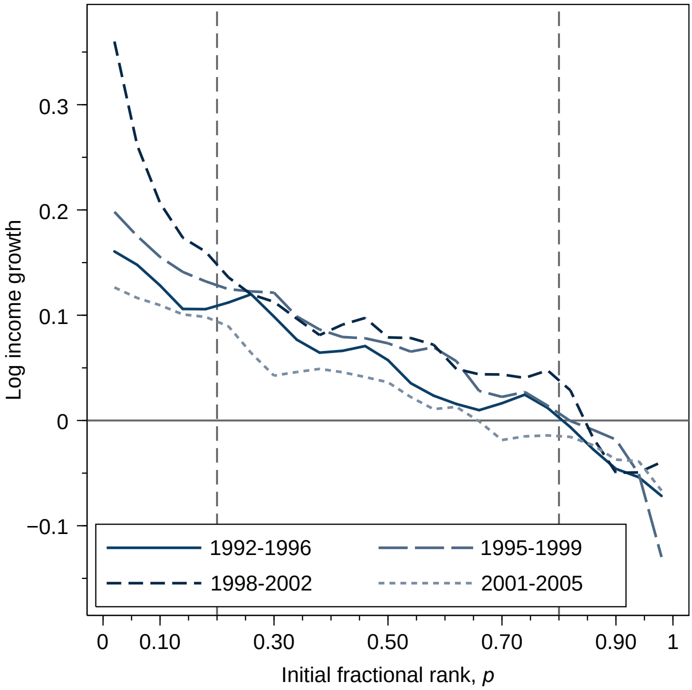


Figure 10.7 Individual income growth and mobility profiles. *Source: Jenkins and Van Kerm (2011).*

[16] As the authors explain, the negative slope to each curve is driven by "regression to the mean," and so the substantive interest is mostly in the changes in slopes of the curves rather than the slopes themselves (as well as their heights).

In sum, we have reviewed a portfolio of tabular and graphical devices for summarizing income mobility between two periods. By standardizing marginal distributions in different ways, different aspects of the mobility process can be focused on, and for individual income growth, there are separate devices.

Within-generation income mobility analysis has tended to use graphical summaries and comparisons rather more than between-generation mobility analysis, which has mainly relied on transition matrix tabulations for detailed summaries of the mobility process. In part, this emphasis is because the mobility concept most associated with intergenerational mobility is pure positional change totally separate from any changes in the marginal distributions. Nonetheless, there do appear to be opportunities forgone to use other methods to describe the distribution.

Our final observation here is that there appear to be no straightforward descriptive summaries that directly highlight the concepts of mobility as longer-term inequality reduction or as income risk. We consider the former case later. In the latter case, one wants something analogous to the mobility profile but, instead, of summarizing expected (average) income growth conditional on base-year income or income position, one would summarize conditional income dispersion.

10.3.2 Mobility Dominance

Dominance checks are a widely used part of the analyst's toolbox for comparing univariate distributions of income. To what extent can and should this be the case for mobility comparisons? We identify three main approaches.

The most well-known dominance results are those of Atkinson and Bourguignon (1982). The results are derived with reference to the social welfare framework discussed earlier. Social welfare is the expected value of individuals' utility-of-income functions defined over period-1 and period-2 income, where individual utility is a concave transformation of the per-period utilities of income and also increasing in each income.

Welfare comparisons of differences in mobility for bivariate distributions f and f^* are based the difference

$$\Delta W = \int_0^{a_y} \int_0^{a_x} U(x, y) \Delta f(x, y) \mathrm{d}x \mathrm{d}y, \tag{10.3}$$

where $\Delta f(x, y) = f - f^*$ is the difference in bivariate densities and the same $U(.)$ is used for the social evaluation of each distribution (cf. Equation 10.3).

Analysis has focused on the case in which the marginal distributions x and y are identical, and SWFs satisfy the conditions $U_1 \geq 0$, $U_2 \geq 0$, and $U_{12} < 0$ (guaranteed if $U(x,y)$ is a concave transformation of the sum of the per-period utilities). Atkinson and Bourguignon (1982) showed that a necessary and sufficient condition for a welfare improvement $\Delta W \geq 0$ is that $\Delta F(x,y) \leq 0$ for all x and y. That is, differences in the cumulative bivariate distribution are lower at each point (a first-order stochastic dominance condition).

What sorts of differences between joint distributions are associated with such conditions being satisfied? Atkinson and Bourguignon (1982) discussed the case of a "correlation-reducing transformation," which leaves the marginal distributions unchanged but reduces the correlation between x and y:

$$\left\{\begin{array}{ccc} & x & x+h \\ y & \text{density reduced by } \eta & \text{density increased by } \eta \\ y+k & \text{density increased by } \eta & \text{density reduced by } \eta \end{array}\right\}, \quad \text{where } \eta, h, k > 0.$$

When the bivariate distribution is represented using a transition matrix, this transformation is equivalent to shifting probability mass away from the matrix diagonal.[17] The cumulative density can be straightforwardly derived by cumulation across cells of the transition matrix starting from the lowest origin and destination group. For comparisons of two transition matrices, first-order welfare dominance exists when the difference in cumulative densities in corresponding cells is everywhere of the same sign. Atkinson (1981a,b) demonstrates the approach in action using intergenerational income data for Britain. Further examples are provided later in this chapter.

The dominance result is a notable addition to the toolbox for comparisons of bivariate distributions but, perhaps surprisingly, has not been widely used. There are several reasons for this. The first is that, although relevant to evaluations of pure positional change mobility, the Atkinson–Bourguignon SWF is primarily sensitive to mobility as reversals rather than mobility as origin dependence (see the earlier discussion).[18]

Second, the first-order dominance checks have not provided clear cut rankings in practice (cf. Atkinson, 1981a,b). A natural reaction in this case is to seek unanimous mobility rankings according to more restricted classes of SWFs using second- and higher-order dominance checks. Atkinson and Bourguignon (1982) provide the theoretical results. The problem, however, is that the additional restrictions on the SWF are hard to interpret. They involve the signs of third- and fourth-order partial derivatives of $U(x,y)$. Although Atkinson and Bourguignon pointed out that in the case of homothetic preferences, "the signs of higher derivatives depend on the relation between the degree of "inequality aversion" … and the degree of substitution" between periods (Atkinson and Bourguignon, 1982, p. 18), i.e., the relation between parameters ε and ρ discussed earlier, they do not elaborate. It is difficult to understand what the sign conditions mean in everyday language.

Third, analysts may be interested in alternative concepts of mobility besides positional change. Individual income growth is the most prominent example of this situation. As

[17] But see also Jenkins (1994) and Fields and Ok (1999a). Both articles question the intuitive attraction of linking correlation-reducing transformations with more mobility if the transformations are made off the diagonal.

[18] Gottschalk and Spolaore (2002) modified the social welfare function but did not derive dominance results.

discussed earlier, researchers have used social evaluation functions that are increasing functions of a measure of "distance" between first and second period incomes for each individual i, $d(x_i,y_i)$ and defined social welfare as the socially weighted sum over individuals of the d_i. For instance, Fields et al. (2002) undertook checks based on comparisons of pairs of CDFs of d_i, where d_i is defined in six different ways in their empirical application. However, as remarked earlier, their SWF has unappealing properties. The challenges involved in the derivation of stochastic dominance results for Fields and Ok (1999b) type measures of nondirectional income movement are discussed by Mitra and Ok (1998). Van Kerm (2006, 2009) explicitly derived dominance results for two classes of SWF defined over the d_i. The first is when the social weights are simply assumed to be positive. Van Kerm showed that unanimous rankings by this evaluation function are equivalent to nonintersections of mobility profiles (the graphical device discussed earlier), a first-order dominance result. If one also assumes that the social weights are nonincreasing functions of base-year income ranks (poorer individuals receive higher weights), unanimous social welfare rankings are equivalent to nonintersections of cumulative mobility profiles. Bourguignon (2011) showed that dominance conditions can be derived for SWFs more closely related to Atkinson and Bourguignon (1982) ones, but the conditions are difficult to interpret intuitively and, in any case, are restricted to the case in which marginal distributions in the initial year are identical.

Dardanoni (1993) derived stochastic dominance results for rankings of mobility processes that are summarized by transition matrices, focusing on pairs of monotone matrices with the same steady-state income distribution.[19] The SWF is defined on a vector containing each individual's lifetime expected utility (the discounted sum of per-period utility values, where each income class has a common utility value associated with it; there is no within-class inequality in utility). Overall social welfare is not the average of the individual lifetime expected utilities because linearity combined with anonymity would imply that mobility is irrelevant for social welfare assessments (as discussed earlier). Instead, Dardanoni's SWF is "a weighted sum of the expected welfares of the individuals, with greater weights to the individuals who start with a lower position in the society" (Dardanoni, 1993, p. 371). Thus there is a direct parallel with the social weight system employed in the welfare function used by Van Kerm (2006, 2009).

Dardanoni shows that unanimous social welfare rankings by this evaluation function can be checked by comparisons of the cumulative sums of the "lifetime exchange" matrices corresponding to the two transition matrices. (A lifetime exchange matrix summarizes

[19] Monotone transition matrices are those in which each row stochastically dominates the row above it. Essentially, being in a higher income class in the initial period means improved prospects in the second period. Most empirically observed transition matrices are monotone or approximately so (Dardanoni, 1993). If a regular transition matrix characterizes a first-order Markov chain, there is a constant long-run steady-state marginal distribution corresponding to that matrix.

the joint probability that an individual starting in some income class i is in lifetime income class j.) These matrices depend on the discount factor underlying them: Although in general mobility processes that improve the position of initially poorer individuals are more highly valued, the timing of utility receipt also matters. Dardanoni (1993) provided additional results for checking the robustness of dominance results to the choice of discount factor. The fact that actual societies may not be in steady state and transition matrices may imply different steady-state distributions limits the applicability of the dominance results. Dardanoni (1993) acknowledged this, but also pointed out that this could be remedied by focusing on bistochastic quantile transition matrices (as Atkinson, 1981a,b) did, in which case attention is restricted to changes in relative position). The orderings derived differ from those of Atkinson (1981a,b), however, because the SWF is different. For instance, Dardanoni (1993) pointed out that maximal mobility according to his ordering corresponds to the situation of origin independence, not rank reversal. Finally, we observe that Dardanoni's dominance results appear to have been rarely used. As with the results of Atkinson and Bourguignon (1982), we suspect that is because applied researchers have found them relatively complicated to interpret and implement.

In sum, we have shown that there are dominance results for mobility comparisons, but the "toolbox" is much less settled than it is for comparisons of univariate income distributions. In part, the reason comes back (again) to the fact that there is a multiplicity of mobility concepts and (related) a lack of consensus about how to specify the SWF function in the bivariate case.

10.3.3 Mobility Indices

In this section, we review indices that might be used to summarize intra- and intergenerational income mobility. After a brief discussion of generic properties of indices, we discuss some commonly used measures of bivariate association—what Atkinson et al. (1992) refer to as "intuitive" measures—and then move on to more specialist indices (i.e., ones more directly corresponding to the various mobility concepts identified earlier). Whether an index focuses on positional change, individual income growth, longer-term inequality reduction, or income risk accounts for many of its properties. There are general features on which we contrast indices.[20]

First, there are different normalizations. Although all indices equal zero in the case in which there is complete immobility, there is no shared maximum mobility value and, indeed, some measures have no maximum value imposed (principally the indices of income growth and income risk). Second, there is a distinction between "pure" measures of positional change and other indices. The former indices, of exchange mobility, are sensitive only to the (re)ordering of individuals and hence with values unaffected by any monotonic transformation of each income between time periods (or, equivalently,

[20] This discussion draws on Jenkins and van Kerm (2009).

also unaffected by changes in the marginal distributions of income). By contrast, structural measures register mobility even if ranks are constant but the income values associated with those positions change over time.

Third, and related, indices differ in how they reflect income changes that are common to all persons, whether by the same proportion or by the same absolute amount. Measures are "strongly relative" ("intertemporally scale invariant") if equiproportionate income growth does not affect the mobility assessment. Measures are "weakly relative" (or "scale invariant") if the units in which income are measured are irrelevant but, by contrast with strongly relative measures, equiproportionate income growth may count as mobility.[21] There are also translation invariance counterparts of these properties. Again, the principal distinction is between measures of pure positional change (exchange mobility)—which satisfy both intertemporal translation and scale invariance—and the other indices. For example, most indices of longer-term inequality reduction are scale invariant but not intertemporal scale invariant. Most indices of individual income growth are neither intertemporal scale nor translation invariant.

Fourth, there is the issue of directionality, which refers to the roles played by the base year and current year in mobility assessments. An index is directional if it matters whether a particular income change refers to a change from a base year to a current year or vice versa. This is relevant if one wishes to take the temporal ordering of changes into account, and this is particularly important for measures of individual income growth, as one would want to treat differently an income change from 100 to 150 and an income change from 150 to 100. One would want the former to represent an improvement in circumstances, and the latter a deterioration.

Fifth, indices may satisfy various decomposability properties. Mobility indices may be (additively) decomposable by population subgroup, as inequality indices are, according to which total mobility can be written as the weighted sum of mobility within subgroups defined by an exhaustive nonoverlapping partition of the population in question according to some characteristic (e.g., sex, age, or education) plus (possibly) a term representing between-group mobility. Most indices of longer-term inequality reduction can be decomposed, thus and so too can individuals of income growth, though there is typically no between-group mobility term in that case.[22] Measures based on changes in ranks are not decomposable because in general there is no one-to-one correspondence between an individual's rank in his/her subgroup and in the population as a whole.

A second type of decomposability is into structural and exchange components. Unlike decompositions by population subgroup, these decompositions are not additive

[21] For more on this distinction, see Fields and Ok (1999a).

[22] Such decompositions have mostly been used to provide anatomies of mobility during a single time period, rather than for accounting for the correlates of changes in mobility between two time periods in terms of the relative importance of changes in subgroup sizes and mobilities and between-group mobility changes.

and rely for their derivation on the use of counterfactual income distributions representing the situations when there is an absence of exchange mobility or of structural mobility. On this, see, e.g., Markandya (1984), Ruiz-Castillo (2004), and especially Van Kerm (2004).

A third decomposition idea, most commonly exploited in measures of individual income growth or income flux refers to intertemporal consistency—whether mobility calculated for income changes between times t and $t+s$ is the sum of the mobility between times t and $t+r$ and between $t+r$ and $t+s$ (with $r<s$) or, alternatively, the product. This is the concept of additive (alternatively, multiplicative) path separability or path independence. A fourth mobility-related decomposition relates changes between two years in *inequality* measured by the change in a generalized Gini coefficient to the sum of two mobility indices—one of progressive income individual income growth and the other of reranking. See Jenkins and Van Kerm (2006) for details.

We refer to these features at several points in what follows. We now turn to consider the most commonly used "statistical" or "intuitive" measures of (im)mobility, which are the Pearson (product moment) correlation, r, between the log of incomes at two time points or its close sibling Beta (β), the slope coefficient from a least-squares linear regression of log(period-2 income) on log(period-1 income):

$$r=\beta\frac{\sigma_1}{\sigma_2}, \tag{10.4}$$

where σ_1 and σ_2 are the standard deviations of log incomes in periods 1 and 2. Put differently, r is β scaled by the changes in inequality in the marginal distributions as assessed by the variance-of-logs inequality index, and it measures the degree of regression to the (geometric) mean in income between periods 1 and 2. $H=1-r$ is the Hart (1976) index of mobility, the properties of which are discussed in detail by Shorrocks (1993) and often used in the intergenerational mobility context. H ranges between -1 and 1, and $H=0$ in the case of complete immobility.

Beta, as we shall discuss later, has been used in almost every empirical study of intergenerational income mobility ($1-\beta$ is an index of mobility). This is perhaps surprising because it is the positional mobility concept that has been of the greatest interest in this context, and yet Beta and r (or H) reflect structural as well exchange mobility. A perfect linear relationship between period-2 and period-1 incomes ($r=1$, $H=0$) is consistent with unchanged ranks but also income growth. It is sometimes argued (see Section 10.5) that r is more suitable than Beta as a measure of income (im)mobility when undertaking cross-national comparisons on the grounds that r controls for differences in marginal distributions. But such controlling is only done to a rather limited extent, because changes in inequality are only one distributional feature (and uses one particular inequality measure to do so). Differences in marginal distributions would be fully controlled, however, were analysts to employ the Spearman rank correlation rather than r (because both marginal distributions would be standard uniform distributions), and this would also have the advantage in the intergenerational context of focusing on positional change. Note also D'Agostino and Dardanoni (2009a) who provided an axiomatic characterization of the Spearman rank correlation as an measure of exchange mobility, thereby taking it beyond being a mere "statistical" index.

A second question regarding Beta and r is why they should be calculated using log incomes rather than incomes. To be sure, Beta is a unit-free measure (an elasticity), but this begs the question of whether we are interested in immobility as the lack of a linear or log-linear relationship.[23]

All in all, there are probably two reasons for the continuing widespread use of Beta and r in the intergenerational mobility literature. The first, as we discuss in Section 10.5, is that various methods to assess the impact of measurement error, and discussions of the relationship between Beta, r, and sibling correlations, rely on properties of regression and moments. The second reason is simply inertia: researchers continue to use Beta because they want to compare their estimates with those of others before them. The main problem with Beta as a measure that intergenerational mobility researchers have noted is its scalar nature rather than more fundamental concerns about the mobility concepts that are reflected in it. Their developments of "vector" measures take us some back toward the more detailed graphical summaries of bivariate distributions discussed earlier.

For example, instead of fitting a single log-log regression, researchers have estimated quantile regressions of period-2 incomes on period-1 incomes (see, e.g., Eide and Showalter, 1999). (The periods refer to offspring and parental generations, rather than years within a generation.) However, it is not immediately clear what the estimates tell us about (im)mobility. At a technical level, the answer is clear. A quantile regression of, say, the 10th percentile of son's income on father's income allows the researcher to express the 10th percentile of son's income as a function of father's income. The quantile regression coefficient on father's log income then measures the elasticity of the particular quantile of son's income with respect to father's income. Differences in estimates across the quantiles tell us how sensitive different parts of the son's distribution conditional on father's income are to small changes in father's income. However, why these marginal changes, measured by slopes of the conditional quantiles, are of interest is not obvious.

One way to interpret the information provided by the vector of quantile regression coefficients is in terms of the full conditional distribution of son's income. A picture that is familiar to most students of regression analysis is the fitted regression line from a regression with one explanatory variable, with the distribution of the error term around that line drawn in a few different levels of the explanatory variable. In the classical case, all those distributions are the same, or at least have the same variance (i.e., the error term

[23] It might also be that log(income) is viewed as a measure of the utility of income, but we have not seen that argument stated explicitly in the income mobility literature.

is homoscedastic). If the distribution of y_2 (for sons), conditional on y_1 (for fathers), is homoscedastic, all estimated quantiles would have the same slope coefficient (save for random error). If the regression slopes are greater for higher percentiles of the son's distribution, it suggests that the conditional variance of son's income may be increasing with father's income.

Comparisons of the quantile regression estimates with the Beta from the loglinear regression can also reveal some further aspects of the distribution of period-2 incomes conditional on period-1 incomes. The log-linear regression line gives the conditional expectation, and the regression slope for the 50th percentile gives the expected median. If we find that for most of the range of father's incomes that the conditional mean for sons is lower than that of the relevant father's income, this suggests that, conditional on father's income, son's income is skewed to the left, rather than skewed to the right (as is usually true for income distributions). One can also use the predicted percentiles for different values of father's income to generate summary distributional statistics for the conditional distribution. For instance, one can derive the (discrete) CDFs for period-2 (son's) income, conditional on a set of period-1 (father's) income percentiles. One could then check, e.g., whether the distributions first-order stochastically dominate each other. (This relates to Monotonicity assumption of Benabou and Ok, 2000.) Any conditional summary statistics can be generated this way, including inequality statistics such as percentile ratios. Individual income growth summaries such as in Figure 10.7 refer to conditional expectations (means) of distributions like these (except that they are typically drawn at different base-year ranks rather than different base-year income levels). In sum, there are close connections between some of the "vector" measures of mobility and the graphical devices discussed earlier. In what follows, we return to focusing on scalar measures.

The second most common type of intuitive measure is an "immobility ratio" (IR). IRs summarize how much clustering there is on (or, sometimes, also around) the leading diagonal of a transition matrix—and hence summarize positional change. For example, for a decile transition matrix, an IR might be defined as the percentage of all persons who remain in the same decile group between the two periods. (A variant would be to calculate the percentage remaining in the same decile group or on either side of it.) Clearly, the IR equals 100% in the complete immobility scenario. If, instead, there is complete independence of origin, the IR for a decile transition matrix is 20% (52% in the variant). An index of mobility can easily be calculated as $1 - \text{IR}$.

Shorrocks (1978b) proposed a mobility index closely related to the IR, a Normalized Trace measure, equal to $[n - \text{trace}(A)]/(n-1)$, where A is the transition matrix with n income classes and trace(A) is the sum of the transition proportions on the leading diagonal of A. This provides a neatly normalized index: with complete immobility, $\text{trace}(A) = n$ and the normalized trace equals 0; with complete origin independence $\text{trace}(A) = 1$, and so the normalized trace equals 1.

By construction, an IR and the Normalized Trace are insensitive to any differences between transition matrices aside from those in the respective diagonals. Bartholomew's (1973) Average Jump index is a positional mobility measure that addresses this aspect. It is equal to the number of income class boundaries crossed by an individual (whether upward or downward), averaged over all individuals (and equal to 0 in the complete immobility case). One feature of the Average Jump index is that it generalizes to the situation when the researcher has individual-level data on incomes rather than simply grouped data (a transition matrix). The index is then the population average of the absolute changes in fractional ranks (i.e., the ranks normalized to range from 0 to 1 rather than from 0 to the population size).

The transition matrix also offers, as a by-product, measures of low and high income persistence defined in terms of rank-order immobility. Mobility matrices, in which class boundaries are defined in real income terms, also do so. For example, if the lowest class boundary in each period is the poverty line, then the mobility matrix shows the proportion of individuals who are poor in a base period who are still poor in some later period (or who escape poverty). And with repeated longitudinal data for multiple periods, it is straightforward to define "survival probabilities"—the chances that a person remaining poor for τ years, where $\tau = 1, 2, \ldots$. One can also define measures of high income persistence analogously (and we report some estimates in the next section).

This way of thinking about low-income persistence provides a link with the more well-known literature on poverty persistence, especially the approach pioneered by Bane and Ellwood (1986) in which consecutive periods spent poor are aggregated into spells (summarizing the total time spent poor). Rather than looking at spells of poverty (or affluence), one may simply count the number of times each person is poor (or rich) over some fixed time horizon and summarize that distribution. Low-income persistence statistics of this nature are published by, e.g., the UK Department for Work and Pensions (Department for Work and Pensions, 2009) and the Statistical Office of the European Communities (Eurostat).

There is also a nascent literature developing indices of poverty persistence that focuses attention on the way in which people's experience of poverty over time is aggregated, and hence how to compare, say, a history of 3 consecutive years in poverty followed by 3 years of nonpoverty, with a history in which the person was poor every second year in the six-year period. Research in this topic can be found in Foster (2009), Gradín et al. (2012), Mendola et al. (2011), Mendola and Busetta (2012), and Porter and Quinn (2012). This literature works with a time horizon of fixed length and summarizes individuals' experiences within that window, ignoring whether poverty spells were already in progress at the beginning of the window or remained in progress at the end of the window. If one wants to derive the shape of the poverty spell distribution in the population (rather than simply the sample), these issues of left- and right-censoring of poverty spell data (which are ubiquitous) need to be accounted for. They are given great attention in the spell-based literature on poverty persistence following Bane and Ellwood (1986), which, on the other hand, ignores longitudinal aggregation issues.

We now turn to a selection of more specialized indices of positional change that have been less commonly used than the ones mentioned so far. The first is the Gini Mobility index of Yitzhaki and Wodon (2005). It is based on the idea of mobility as (lack of) correlation but, instead of using the Pearson or Spearman correlations, it uses the Gini correlation, which, like other Gini-based measures, focuses on ranks rather than income levels per se. The Gini correlation between the income distributions in periods 1 and 2 is

$$\Gamma_{12} = \frac{\mathrm{cov}(y_1/\mu_1, F_2)}{\mathrm{cov}(y_1/\mu_1, F_1)}, \tag{10.5}$$

where y_1/μ_1 is period-1 relative income (i.e., income divided by the period-specific mean income), F_1 and F_2 are the fractional ranks in the two periods, and cov(.) means covariance. Because $1 - \Gamma_{12}$ is a directional measure of mobility ($\Gamma_{12} \neq \Gamma_{21}$ in general), the overall Gini mobility index is defined as a weighted average of the two possible directional measures, where the weights depend on the inequalities in each marginal distribution, measured using the Gini coefficient (G). That is,

$$\text{Gini mobility index} = \frac{G_1(1-\Gamma_{12}) + G_2(1-\Gamma_{21})}{G_1 + G_2}. \tag{10.6}$$

Yitzhaki and Wodon (2005) showed that if there is no positional change, the Gini mobility index equals 0, it equals 1 if there is complete origin independence, and it equals 2 if there is complete rank reversal.[24]

The Gini mobility index uses a particular weighting function when aggregating changes in individuals' ranks but one that is not immediately clear. By contrast, the King (1983) index takes an explicit welfarist approach in which differences in social weights across ranks are defined and tuned parametrically. The basic building block is the "scaled order statistic" for each individual i, s_i, equal to the absolute magnitude of the difference between i's period-2 income and the period-2 income that i would have had, were she/he to have maintained the same rank in period 1, all expressed relative to mean period-2 income. There is complete immobility if $s_i = 0$ for all individuals. Using an approach analogous to that of Atkinson (1970), King defines his mobility index as the proportion of total period-2 income that society would be prepared to forego to have the mobility observed rather than complete immobility (positional change is socially valued). Assuming a homothetic form for the SWF leads to a mobility index depending on two parameters—the degree of aversion to period-2 income inequality and the degree of aversion to income immobility (larger values of which give greater social weight to

[24] The index of reranking used in Jenkins and van Kerm's (2006) decomposition of inequality change into reranking and income growth components is a directional Gini correlation.

mobility, other things being equal). For generalizations of and commentary on King's approach, see Chakravarty (1984) and Jenkins (1994).

On the one hand, the systematic welfarist approach used by the King index (and others like it) has much to recommend it. On the other hand, it relies on a rather special characterization of what counts as mobility at the individual level (the scaled order statistic), in which the implications for social welfare of changes in ranks are summarized by income values and a particular no-mobility thought experiment. Also the SWF does not depend directly on incomes in period 1, except insofar as they characterize s_i. Compare this with the Atkinson and Bourguignon (1982) SWF defined over incomes in periods 1 and 2 that was discussed earlier. Gottschalk and Spolaore (2002) embraced (and extended) the latter in the first part of their article, but when they later defined specific mobility indices, they use an approach that is similar to King's (1983) in that the social gains from mobility are all expressed relative to a complete immobility reference point, and this is defined in the same way as in the King index. To develop their indices, Gottschalk and Spolaore (2002) also assumed homotheticity in their SWF, and the resulting class has three parameters, representing aversion to multiperiod inequality, rank reversal, and origin dependence. Although each parameter has a clear interpretation when taken individually, thinking about the implications of different combinations of values is more complicated. We are aware of no use of the Gottschalk and Spolaore (2002) indices other than by the authors themselves.

We move now to consider measures of individual income growth. As mentioned in the Introduction, these incorporate two basic ideas: (i) income increases for an individual count positively in the social calculus and income decreases count negatively, and (ii) total income growth is a function of income growth values for each individual (and the measure of each person's income growth depends only on their incomes in the two periods, and not the incomes of other people). The first idea refers to the directionality of the income growth measure. The second is a form of decomposability property across individuals, and also leads to aggregate measures that are decomposable by population subgroups. Although the empirical applications of these measures have all been to intragenerational income mobility, the indices could also be applied to intergenerational income mobility when there is interest in structural mobility over and above exchange mobility.

Fields and Ok (1999b) provided the most well-known aggregate measure of directional income growth in this tradition.[25] They show that directional measures of individual income growth that satisfy the properties of scale invariance, subgroup decomposability, and multiplicative path separability must take the form

[25] We review their nondirectional indices of income "flux" later.

$$D1 = c\left[\frac{1}{N}\sum_{i=1}^{N}\left(\log(y_i) - \log(x_i)\right)\right], \tag{10.7}$$

where c is a normalizing constant, which may be set equal to one, and N is the population size. That is, overall income growth is the average of individuals' proportional income growth. This is the case in which (directional) distance between incomes, $d(x_i,y_i) = \log(y_i) - \log(x_i)$. Observe that the social weighting scheme treats all individuals the same, regardless of their base-year income and regardless of how much income growth each experiences. Both these aspects, and some other generalizations, have been incorporated in later work.

Demuynck and Van de gaer (2012) used axioms similar to Fields and Ok (1999b), but explored the implications of assuming additive as well as multiplicative path separability, and also of imposing an axiom of "priority for lower growth" that builds in aversion toward inequality in the individual growth rates. The axiom states that "aggregate growth increases more when additional income growth is allocated to individuals with lower income growth than when it is allocated to individuals with higher income growth" (Demuynck and Van de gaer, 2012, p. 750). The authors prove that the measure satisfying their axioms is of the form:

$$S = \frac{1}{N^{\delta}}\sum_{i=1}^{N}\left(i^{\delta} - (i-1)^{\delta}\right)\widetilde{d}_i, \quad \text{with } \delta \geq 1. \tag{10.8}$$

Given a measure of individual-level income growth for each person, d_i, $\widetilde{\mathbf{d}}$ is the vector of such income "distances" ordered from largest to smallest. If multiplicative path separability is among the axioms, then $d_i = (y_i/x_i)^{\pi}$ or if, instead, additive path separability is assumed, then $d_i = \pi(\log(y_i) - \log(x_i))$, with $\pi > 0$ in both cases.

When $\delta = 1$, the general indices reduce, in the first case, to the directional measure of Schluter and Van de gaer (2011) and, in the second case, to the Fields and Ok (1999b) measure described earlier (with $\pi = c$, and also normalized to 1). In Schluter and Van de gaer's (2011) index, π is a sensitivity parameter, with higher values increasing the "distance" measured between incomes in period-1 and period-2 but keeping ranks the same. Demuynck and Van de gaer (2012, p. 754) remarked that when $\delta = 1$, correlation-reducing transformations to incomes in either period of the kind discussed earlier increase mobility according to S but $D1$ is insensitive to such changes. In the more general case, with $\delta > 1$, more weight is given to individuals with smaller values of d_i. When $\delta = 2$, the weights are like the weights used to characterize the Gini coefficient of inequality and when $\delta \to \infty$, only the smallest d_i counts. In these more general cases, S is no longer additively decomposable by population subgroup, and it is possible for correlation-decreasing transformations to reduce mobility. The larger question, however, concerns the social desirability of "priority for lower growth": why should we be concerned about the inequality of individual growth rates (the d_i) independently of incomes in the initial or final period? Because of this issue, and (related) the greater complexities involved with using a two-parameter index, we conjecture that empirical researchers will be more likely to use S with $\delta=1$ than the more general case.

The directional measures of income growth of Jenkins and van Kerm (2011) are built using a different approach and relate to a SWF defined as the weighted average of the d_i (see Section 10.2), in which the social weights are a decreasing function of period-1 income ranks, defined using a single-parameter generalized Gini scheme.[26] Put differently, Jenkins and van Kerm (2011) build in a social preference for pro-poor income growth, and the choice of different parameter values provides indices ranging from limiting cases in which aggregate growth is the simple average of the d_i values (as with $D1$) or in which only the growth rate for the poorest period in the initial year counts.[27] Palmisano and Van de gaer (2013) provide an axiomatic characterization of the Jenkins and van Kerm (2011) class of measures. The usefulness of these indices rests largely on the extent to which the concept of pro-poor income is viewed as a desirable normative principle: see the discussion in Section 10.2 about the link between progressive income growth and inequality reduction.

The pioneering paper on mobility as reduction in the inequality of longer-term income is by Shorrocks (1978a). The essential insight is that, were one to longitudinally average each person's income over a number of years (T, say), the inequality in these averaged incomes would be less than average annual inequality because each individual's income fluctuations would be smoothed out and no longer contribute to aggregate cross-sectional dispersion in incomes for the T-year accounting period. Shorrocks (1978a) defined a measure of income rigidity, $R(T)$, equal to the ratio of inequality among T-averaged incomes ("longer-term" inequality) to the weighted average of single-year inequality values:

$$R(T)=\frac{I[Y(T)]}{\sum_{k=1}^{k=T} w_k I[Y^k]}. \tag{10.9}$$

$I[Y(T)]$ is the inequality in T-averaged incomes, and $I[Y^k]$ is inequality in period-k incomes calculated using the same inequality index (e.g., $I[Y^1]$ is inequality in period-1 incomes). The weights w_k are the proportion of aggregate T-averaged income received in period k (i.e., $w_k=\mu_k/\mu$), and the weights sum to unity. Shorrocks shows that if one restricts attention to conventional relative inequality indices, then R is bounded

[26] This scheme is like Demuynck and Van de gaer's (2012), except that the weights are applied to period-1 ranks, and not to d_i values.

[27] Jenkins and Van Kerm's (2011) classes of measures focus on the cases in which income growth rates are defined in proportional or absolute terms (i.e., $d_i = \log(y_i) - \log(x_i)$ or $d_i = y_i - x_i$).

above by 1. When there is complete rigidity in relative incomes, inequality in each period corresponds to inequality for the longer accounting period.[28] The more frequent or larger that income changes are, the less rigid the income system, and thus one may define a measure of mobility: $M(T) = 1 - R(T)$.

As Shorrocks (1978a, p. 178) puts it, "mobility is regarded as the degree to which equalization occurs as the observation period is extended." In terms of the properties discussed earlier, $M(T)$ is a nondirectional index and scale invariant (because it is defined in terms of relative incomes), but not intertemporal scale invariant (given the way in which the per-period weights are defined). Although R and M are usually used to describe within-generation mobility, in principle they could also be used to describe mobility between generations. R and M are distinctive in that they are well defined when there are data for many periods, but they can also be calculated if there are only two (the typical situation with intergenerational data).

A nice feature of the Shorrocks approach is that it can be used in two ways. The first is to calculate a single index value conditional on a particular value of T (and inequality index). This fixed-window calculation can be employed, e.g., to examine trends in income mobility over time in a country using moving fixed-width windows. Second, one can examine how $R(T)$ changes as T is increased from its minimum value of 1 to some larger maximum (i.e., there is one window, the width of which is varied). The resulting rigidity and mobility profiles provide a straightforward graphical device for comparisons of the extent of mobility within a country, and also comparisons across population subgroups and countries. Rigidity profiles for the United States and Western Germany from a pioneering cross-national study of income mobility discussed further in the next section are shown in Figure 10.8. The profile for Western Germany lies everywhere below that for the United States: Whatever the accounting period used, mobility is greater in Western Germany than in the United States.

Clearly, the values derived for the Shorrocks indices are conditional on the inequality index employed for the calculations. It is also well known that inequality indices differ in the sensitivity to income differences in different parts of the income distribution (Atkinson, 1970). So, it is important to know how estimates of rigidity and mobility relate to choice of inequality index, and how differences in inequality index sensitivity translate into mobility index sensitivity. It has been found as an empirical regularity, from Shorrocks (1981) onward, that using different indices can make a big difference to the estimates of R derived and also that the Gini coefficient tends to show greater R values

[28] By conventional relative inequality indices we mean all those that are convex functions of relative incomes (incomes expressed relative to the mean income), that is, all those that satisfy the Principle of Transfers. This excludes indices such as the variance of log incomes. R is bounded below by zero, assuming all incomes are positive.

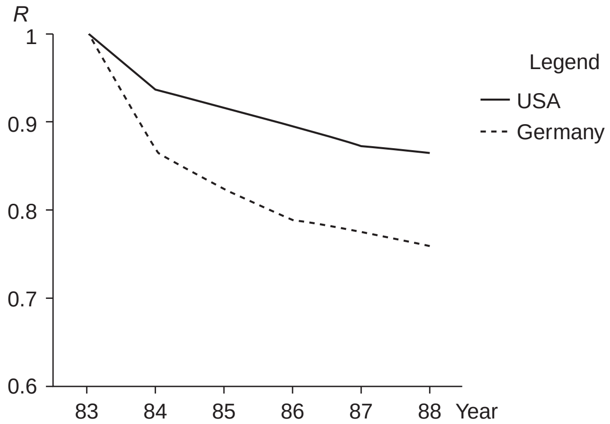


Figure 10.8 Income rigidity (longer-term inequality expressed as a fraction of total inequality) falls as the time period is lengthened. *Note: Income is posttax posttransfer income. The Shorrocks rigidity index R is computed using the Theil index of inequality. "Germany" refers to the federal states of Western Germany. Source: Burkhauser and Poupore (1997, Figure 2).*

than other inequality indices. The explanation is that "[since] the main effect of cumulating income is to average out incomes that are temporarily high or low, the strongest egalitarian trend will be found in the tails. The distribution of relative incomes in the middle range is not substantially affected by cumulating incomes over time" (Shorrocks, 1981, p. 182). Combine this information with the fact that the Gini coefficient is relatively insensitive to income transfers in the tails of the income distribution, and we have the result.

The sensitivity of *R* to the choice of inequality index is examined more systematically by Schluter and Trede (2003). For the two-period case, they show that the global rigidity measure, *R*, can be expressed, to a good approximation, as the weighted average of "local" rigidity comparisons at each point along the income range of a value for the longer-term averaged income, and the average of the per-period distributions. Differences in global measures arise, therefore, from a combination of differences in the way the different inequality indices summarize local comparisons at each point along the income range, and the different weighting systems that they incorporate. Schluter and Trede (2003) show that the sensitivity of mobility measure to choice of inequality index is partly dependent on data, but they also show some clear empirical regularities. For example, the weighting functions for commonly used generalized entropy indices and the Gini coefficient are broadly similar around the middle of the distribution (relative income = 1) and tend to place greater weight on mobility at the tails of the distribution. In addition, the overall U-shape for the weighting function is distinctly shallower for the Gini than for the other indices (as Shorrocks argued). Given the ready availability of longitudinal data on incomes nowadays (see the next two sections), it is straightforward for researchers to examine sensitivity empirically.

Refinements to the Shorrocks approach have gone in two main directions. The first addresses the assumption that individuals are able to smooth incomes across time: see the discussion in Section 10.2. This aspect is relaxed by Maasoumi and Zandvakili (1990) and Zandvakili (1992), building on Maasoumi and Zandvakili (1986). See also the survey by Maasoumi (1998). The basic idea is to allow for different degrees of substitutability between incomes in different periods. Thus, rather than defining longer-term income for each individual as the simple arithmetic average, it is defined as a generalized mean for which the choice of a parameter tunes the degree of substitutability. A common parameter is used for each individual, and yet one would expect the ability to smooth income over time to vary with, e.g., income level. Incorporating such heterogeneity into an index would be a rather complicated exercise and has not been done, as far as we are aware. As it is, researchers wishing to implement the Maasooumi–Zandavakili variant on R need to choose a substitutability parameter as well as inequality index. Of course, estimates can easily be derived for a number of combinations but the volume of results produced is probably one reason the approach is not commonly used. Also, the empirical illustrations provided by Maasoumi and Zandvakili (1990) and Zandvakili (1992) tend to suggest that the more general index tended to provided qualitatively similar results to that of the Shorrocks approach.[29]

The second refinement to the Shorrocks approach is to reconsider the reference point against which longer-term inequality values are compared. The main argument of Fields (2010, p. 410) is that "[w]hat we as empirical researchers would want to know in a given context is the extent is the extent to which the mobility that takes place works to *equalize* longer-term incomes relative to base, *disequalizes* longer-term incomes relative to base, or has no effect" (emphasis in original). This leads to Fields' proposal that the denominator in the expression for R be changed from the weighted average of the per-period inequalities to the inequality in first-period income. Chakravarty et al. (1985) emphasized rather different aspects in the derivation of their mobility index: they were concerned with "ethical" indices of relative income mobility, which are derived from SWFs and measure changes in welfare. Mobility is the percentage change in social welfare (measured by the equally distributed equivalent income, defined in the Atkinson (1970) sense) of the actual distribution of longitudinally averaged incomes compared to what social welfare would have been in the completely immobile benchmark distribution—taken to be the observed period-1 distribution.

[29] Maasoumi and Zandvakili (1990) and Zandvakili (1992) were also the first to provide decompositions of total mobility calculated using Shorrocks-Maasoumi-Zandvakili indices into components representing within-group mobility and between-group mobility. They did not provide formulas for the decomposition, however. For these, see Buchinsky and Hunt (1999, p. 354).

If the same welfare function is used to evaluate both distributions, and that SWF is homothetic, then the mobility measure "has a natural interpretation; it is the percentage change in equality of the aggregate distribution compared with the first-period benchmark" (Chakravarty et al., 1985, p. 6). Although the authors go on to state that it appears there is no convincing ethical argument for applying the same welfare function to both distributions, all empirical applications that we are aware of have applied the same welfare function. The class of mobility indices for the two-period case is then defined as (Chakravarty et al., 1985, p. 8):

$$C = \frac{1 - I[Y(T)]}{1 - I[Y^1]} - 1, \tag{10.10}$$

where I is a relative inequality index equal to one minus an index of relative equality (as is the case with the Atkinson (1970) class of inequality indices). It turns out that the Fields (2010) mobility index, $1-[I[Y(T)=I[Y^1]$, equals κC where $\kappa=(1-I[Y(T)])=I[Y^1]$, and so the measures are closely related (assuming the same inequality index is applied in each case). But it is possible for them to differ about whether mobility has increased or not: the value of κ matters. In short, ethical index C always evaluates mobility as welfare-increasing (but of different degrees), whereas the more descriptive Fields (2010) index allows mobility to be positive or negative. A more fundamental issue, common to both indices, is whether one agrees with the proposal to accord special normative status to period-1 incomes relative to incomes in other periods—which is an issue that has arisen with other mobility measures as well.

The concept of comparing short- and longer-term incomes has been used to examine poverty persistence in particular as well as income mobility in general. The basic building block is again "longer-term income," a measure of longitudinally averaged income for each individual, and people are defined as "chronically" poor if their longer-term income is less than the poverty line. Chronic poverty in aggregate is the poverty in the population calculated using a poverty index that is additively decomposable over people and time (e.g., a member of the Foster et al., 1984 class). Transitory poverty is Total Poverty (poverty calculated over individuals and separate time periods) minus Chronic Poverty. The main papers to date in this tradition are Rodgers and Rodgers (1993, 2009), Chadhuri and Ravallion (1994), and Jalan and Ravallion (1998). See also the development by Duclos et al. (2010), which takes a more explicitly welfarist approach. As with the Shorrocks mobility measures, there is an important issue concerning how longer-term incomes are calculated and (related) the assumptions made about abilities to income smooth. See, e.g., the discussion by Rodgers and Rodgers (1993, pp. 34–35).

The final group of more specialist mobility indices we discuss are those that summarize notions of income risk. These can be classified in two main ways. On the one hand, there are measures of the transitory variance of (log) income, calculated using either model-based or nonparametric approaches and generally requiring income data for multiple periods. On the other hand, there are measures of income flux, income movement, and volatility, generally defined over incomes in two periods only. We consider the approaches in turn and discuss the relationships between them and the measures of longer-term inequality reduction.

To fix ideas,[30] suppose that the dynamics of income for each individual can be described using the canonical random effects model

$$\log y_{it} = u_i + v_{it}, \tag{10.11}$$

where y_{it} now refers to the income for person i in year t. It consists of a fixed "permanent" random individual-specific component, u_i, with mean zero and constant variance σ_u^2 (common to all individuals), and a year-specific idiosyncratic random component with mean zero and variance σ_v^2 (common to all individuals) that is uncorrelated with u_i. Thus total inequality as measured by variance of log incomes is equal to the sum of the variance of "permanent" individual differences plus the variance of "transitory" shocks:

$$\sigma_t^2 = \sigma_u^2 + \sigma_v^2. \tag{10.12}$$

Assuming that permanent differences are relatively fixed over time, changes over time in income inequality (σ_t^2) arise mostly through changes in the variance of the transitory component. The interpretation of this latter component as idiosyncratic unpredictable income change leads to the association of changes in its variance with changes in income risk.

This canonical model is patently unrealistic in several respects, and three types of extension have been incorporated.[31] The first additional factor allows the relative importance for overall inequality of the permanent and transitory components to change with calendar time. For example, if there is an increase in the demand for skilled labor, and permanent component of income represents relatively fixed personal characteristics related to skills (for example human capitals of various kinds), then greater inequality resulting from widening differences over time in returns to skilled versus unskilled labor can be represented as the growing importance of the permanent component. In contrast, a secular trend toward greater labor market flexibility can be represented as a growth in the importance of transitory variations. The second additional feature is persistence in

[30] The exposition in the next few paragraphs draws heavily on Jenkins (2011a, chapter 6).

[31] For surveys of model specification and estimation methods, see inter alia, Baker and Solon (2003), Guvenen (2009), Meghir and Pistaferri (2011), Haider (2001), Shin and Solon (2011), Moffitt and Gottschalk (2012), and the references cited in these sources and in Section 10.4. Note that these variance components models have usually been applied to data on men's earnings and only rarely to household income. See Jenkins (2011a, chapter 6) for further discussion.

transitory shocks. The factors leading to a temporary fall (or rise) in income in one year are likely to have effects that last longer than a year: a transitory shock persists but with diminishing impact and eventually dies out. An example might be an accidental injury leading to a reduction in work hours that diminishes over time. This is usually characterized using an autoregressive moving average process for v_{it}.

The third modification to the canonical model is to allow the fixed individual component to change over time. Two main approaches have been followed, originally distinct but now commonly combined. One is to allow u_i to vary over time via a "random walk": this year's value is equal to last year's value plus or minus a random element. The second approach allows for individual-specific rates of growth in income (the "random growth" model). The expression for the permanent component is modified so that it also varies linearly with time but with heterogeneity in this slope. Both a random walk and random growth lead to a fanning out of the income distribution over time, other things being equal. Rankings are preserved; those at the bottom stay at the bottom but fall further behind those at the top, who stay at the top. It is increases in the transitory variance that increase mobility in the sense of reranking.

The estimation of transitory variances (mobility) and permanent variances using these models is common, but has also been criticized on the grounds that estimates are sensitive to the particular model specification employed, and there are potential identification issues with the relatively short household panels used to estimate the models (see, e.g., Doris et al., 2013; Guvenen, 2009; Shin and Solon, 2011). This has led to simpler nonparametric methods also being regularly used.

The most common nonparametric method for deriving estimates of variance components is the window-averaging method first employed by Gottschalk and Moffitt (1994), also known as their BPEA method (the acronym refers to the journal in which their work was published). The BPEA method works by first calculating the longitudinal average of each person's log income over a time window of fixed width, say T years. This provides an estimate of the person's "permanent" income for that period and is directly analogous to the longer-term income concept used to derive R except that it refers to averaging of log incomes. (If Equation 10.11 describes the income-generation process, the longitudinal average is an estimate of u_i.) The transitory incomes for each individual within the window are derived as a difference between this permanent income and observed log income, from which can be calculated the individual-specific transitory variance. The overall sample transitory variance is the average of these variances. The sample permanent variance for each window is calculated from the differences between each person's permanent income and the sample grand mean of these, with an adjustment to account for the fact that the mean contains a proportion of the transitory component that has not been fully averaged to zero over the T-year window. See Gottschalk and Moffitt (2009, p. 7) for full details of the formula, and Kopczuk et al. (2010, p. 98) for a small variation on the same theme. The BPEA method is known to provide biased estimates of the transitory variance and its trend if the permanent component's contribution changes over time (see, e.g., Shin and Solon, 2011). Using shorter-width windows for the calculations (smaller T) reduces the potential impact of this problem, but at the cost of reducing the statistical reliability of the estimate of each person's permanent income.

It is inevitable that measures derived using methods like the BPEA one will reflect the variability from permanent shocks and not only from transitory shocks. Shin and Solon (2011, p. 977) argued that this is a virtue of such measures: "The recent interest in volatility trends stems in large part from a concern about whether earnings risk has increased. Because permanent shocks, such as those experienced by many displaced workers, are even more consequential than transitory ones, it makes good sense to include them in the measurement of earnings volatility." Their own calculations use instead a measure of volatility that will be discussed shortly.

Both of the two main methods for estimating transitory variances have potential weaknesses, and there are virtues in using both as well as other measures (such as of volatility) as a sensitivity check. (This is increasingly done, as the next section shows.) Regardless of estimation method, there is a distinction between measures of mobility that are based on the transitory variance itself and measures that are based on the transitory (or permanent) variance expressed as a proportion of the total variance. Most discussion uses the former as the definition of mobility in the form of income risk.

Some authors also present estimates of the permanent variance expressed as a proportion of the permanent variance, and note that, if estimated using the BPEA method, there is a close relationship with the estimates of the Shorrocks measure of income rigidity R. See, e.g., Burkhauser and Couch (2009) and also Chen and Couch (2013, p. 202), who state that they prove that "under one testable condition a measure of economic mobility formed by the ratio of permanent to total variance employing the methods of Gottschalk and Moffitt (1994) is equivalent to the Shorrocks R constructed with a Theil General[ized] Entropy Index." It is clear that there must be some relationship, but we believe that it is not as close as stated by these authors, for the simple reason that the BPEA method calculation uses log incomes, and calculations of R invariably use incomes expressed in levels rather than logs. Evidence showing that a BPEA-estimated ratio of permanent to total variance and Theil-based estimate of R can move in opposite directions appears in Bayaz-Ozturk et al. (2014, Figure 2). For a related discussion, see also Shorrocks (1981, Section 6), who considers the shape of the profile for $M(T)$ in the case in which incomes—not log incomes—follow the basic canonical random effects model (cf. Equation 10.11) and inequality is calculated using half the squared coefficient of variation. He shows that were the model to hold, $M(T)$, would converge to its limiting value fairly rapidly. Slow convergence is evidence that the canonical model is inappropriate.

Income volatility in a given year t, V_t, is commonly measured by the standard deviation (SD) of the distribution of individual changes in log income between 1 year and an earlier year:[32]

$$V_t = \mathrm{SD}[\log(y_{it+\tau}) - \log(y_{it})]. \tag{10.13}$$

Changes are typically measured over a one- or two-year horizon: $\tau = 1$ or 2. The Fields and Ok (1999b) index of individual income growth ($D1$ discussed earlier) is the mean of the distribution of log-income changes. Volatility is therefore a measure of dispersion of the same distribution using one specific index of inequality. There are further connections: If the Gottschalk–Moffitt BPEA method is used to calculate the transitory variance in the two-period case, the resulting estimate is equal to one-quarter of the variance of the change in earnings (i.e., V_T^2 with $T = 2$). See Moffitt and Gottschalk (2012, p. 218), who also point out that the relationship no longer holds if the data window is longer than two periods.

This brings us to the measures of income flux, most commonly associated with the names of Fields and Ok (1996, 1999b), who proposed a number of measures of nondirectional income movement for the two-period case. Although such indices are rarely related to the measure of income risk discussed so far, their inventors had this application in mind: "A measure of income movement ... identifies how unstable the incomes of individuals have been throughout the time period. Since income instability may cause economic insecurity, ... measure of income movement are useful complements to the traditional measures of relative income mobility" (Fields and Ok, 1999b, p. 455). In their 1996 paper, Fields and Ok consider what they label absolute measures of income movement. First, they propose a number of axioms to describe measures for a fixed population of N individuals: linear homogeneity (equi-proportionate increases in all incomes, in base and final year, lead to the same proportionate increase in the measure); translation invariance; a normalization axiom; decomposability (total mobility for the N individuals is a symmetric function of the income changes for each individual); growth sensitivity (if two bivariate distributions are identical except that in one distribution an individual experiences more income movement than in the other distribution, total mobility differs in the two distributions); and, finally, the axiom of individualistic contribution (the contribution of each individual's mobility to total mobility does not depend on how other people's incomes change).

Fields and Ok (1996) prove that the measure satisfying these seven axioms is the sum over the N individuals of the absolute differences between period-1 and period-2 incomes (i.e., $|y_i - x_i|$, for each individual $i = 1, \ldots, N$). The final step is to consider

[32] See, e.g., Shin and Solon (2011). Other variants use a different definition of proportional income change, most often the arc percentage change See, e.g., Dynan et al. (2012). This has the advantage of allowing for zero income values in the estimation of volatility.

versions of these measures that would enable comparisons across populations of different sizes. Specifically, their per capita measure of absolute measure of absolute income movement is

$$D2 = \frac{1}{N}\sum_{i=1}^{N} |y_i - x_i|. \tag{10.14}$$

Their "percentage" measure is the same as $D2$ except that the denominator is total income in period 1 rather than population size and has been less commonly used perhaps because it is unclear that the base year should be used as the reference point (see our earlier discussion).

In their 1999 article, Fields and Ok took a similar similar set of axioms but also considered scale-invariant measures of movement as well as translation-invariant ones. This leads to the per-capita relative movement index, given by

$$D3 = \frac{1}{N}\sum_{i=1}^{N} |\log(y_i) - \log(x_i)|. \tag{10.15}$$

Both $D2$ and $D3$ are additively decomposable by population subgroup: total income movement can be expressed as the weighted sum of the movement within each subgroup, where the weights are the subgroup population shares. Fields and Ok (1996, 1999b) show that $D2$ and $D3$ also satisfy a different sort of decomposition: in each case, aggregate income movement can be expressed as the sum of a component representing income "growth" for individuals and a residual component that can be interpreted as income "transfers" between individuals. (Slightly different versions of the decomposition apply depending on whether the average of the first component is positive or negative.) It turns out in the case of $D3$ that the growth component of this decomposition is the directional measure of proportionate income growth ($D1$) discussed earlier.

To return to the remarks earlier about the links between measures of income flux and other measures of income risk, observe that the variance of log-income changes between two periods can be written as $E(d_i)^2 - E^2(d_i)$, where E is the expectation operator, and $d_i = \log(y_{it+\tau}) - \log(y_{it})$. That is, volatility-squared is equal to the average of the squared log-income changes, minus the square of the average log-income change. The first term is a measure of income flux in which the distance concept used to record income changes is Euclidean distance. Thus, there is a close relationship between orderings by this measure and a volatility measure when average log-income changes are "small." This Euclidean distance measure is characterized axiomatically by D'Agostino and Dardanoni (2009b), who also compare their approach with that of Fields and Ok (1996, 1999b). A measure incorporating the Euclidean distance concept is also characterized axiomatically by another pioneering paper on the measurement of income movement, by Cowell (1985) (see also Cowell and Flachaire, 2011) The axiom set is rather different in Cowell's

(1985) paper, however, and also leads to parametric classes of subgroup decomposable measures of "distributional change." These indices have rarely been used in empirical applications, however, perhaps because their properties (in particular the implications of choosing different parameters) are rather opaque in comparison with the overt transparency of measures like $D2$ and $D3$.

This completes our review of the many measures of income risk. A question that could be asked about all of them is whether they actually measure income "risk" in a more fundamental sense, namely the ex ante uncertainty aspect drawn attention to by, e.g., Gottschalk and Spolaore (2002). As Creedy et al. (2013, p. 236) remind us, this requires a model of expectations formation based on observed income dynamics. There are also additional complications for welfare evaluations such as the extent to which observed income changes reflect voluntary decisions of individual and families and the extent to which these are insurable (and how these aspects differ across people). These complicated underpinnings are absent from the measures we have discussed. At the other extreme are more structural models such as proposed by Blundell et al. (2008) and Cunha et al. (2005). Our overall assessment is that the measures we have discussed are useful descriptive measures despite these flaws. Their relative simplicity facilitates transparency and interpretation as well as empirical implementation. But they should be interpreted cautiously.

Our final remarks concern the applicability of the mobility measures to intragenerational and intergenerational data. As we have noted, different mobility concepts may be more relevant in one context than another. For example, positional mobility concerns appear of particular relevance to discussions of intergenerational mobility and income growth of particular relevance to discussions of intragenerational mobility. But structural intergenerational mobility is also of interest, and so too is the identification of intragenerational reranking along with income growth. By providing a unified treatment of mobility measures, we hope that some cross-context fertilization may be facilitated. In principle (and data permitting), all the measures we have discussed could be used in either context. In the two sections that follow, reviewing empirical evidence about intra- and intergenerational mobility, we reveal which measures have been used to date.

10.4. INTRAGENERATIONAL MOBILITY: EVIDENCE

This section assesses evidence about within-generational income mobility. It first considers definitional issues, the nature of the longitudinal data available, and issues of empirical implementation, and then it turns to the evidence itself. Our review of the topics is selective. We draw on and refer readers to Jenkins (2011a, chapters 2 and 3) for a much more extensive discussion of data sources for within-generation mobility and related empirical issues, as well as extensive references to other literature. Our survey of evidence concentrates on findings emerging over the last two decades, and gives greatest attention to the United States, with examination of trends over time and cross-national comparisons between the United States and (Western) Germany, but studies for other countries are also considered. Our focus reflects the emphasis in research to date, and this, in turn, is related to the availability of suitable data (as we explain). Also, to make the review manageable, the focus is on mobility of household income rather than of individual labor earnings (though selected earnings studies are referred to). We show how conclusions about trends over time and cross-national differences vary with the mobility concept chosen.

Issues of statistical inference are ignored here. On these, see, e.g., Biewen (2002) and Chapter 7.

10.4.1 Data and Issues of Empirical Implementation

Any study of income mobility faces three "W" issues: mobility of What, among Whom, and When? Studies of trends over time or across countries add another issue, that of comparability. The choices that researchers can make under these headings are much constrained by the sources of longitudinal data that are available. But the data situation has improved substantially over the last two decades. (Contrast the situation described later with the discussion by Atkinson et al., 1992, chapter 3, which focuses on earnings.) Although many of the "W" issues arise in any study of income distribution, looking at mobility adds some extras twists to those arising in cross-sectional analysis.

Mobility of "What" refers to which income sources are included in the definition of "income." Definitions typically range from measures with only a single source (typically earnings from employment) to a broader measure such as household income, which includes multiple sources. Many variations are possible (e.g., labor earnings may refer to employment earnings only, or earnings from all jobs that an individual has, and may also include self-employment earnings, thought often not). There are multiple definitions of income as well. The most common distinction in empirical work is between measures of pretax pretransfer income, pretax posttransfer income, and posttax posttransfer (also often labeled original or market or pregovernment income; gross income; and net, disposable, or postgovernment income, respectively). Pregovernment income typically includes labor earnings, income from savings and investments, and transfers received from nongovernment sources. Taxes usually refer to taxes on income (typically at national level, sometimes also including local taxes) and contributions levied for public pensions. "Transfers" usually refer to cash benefits received from the state.[33]

Mobility among "Whom" refers to the definition of the income-receiving unit. Clearly this is closely related to the issue of What. For example, it is individuals that receive labor earnings. Benefits are assessed and income taxes levied on families and

[33] For a comprehensive discussion of the various definitions and recommendations for measurement, see Expert Group on Household Income Statistics (The Canberra Group, 2001).

households. Individuals not in paid work such as stay-at-home mothers or children, often do not receive income in their own right, but benefit from income sharing with families and households. Putting things another way, note that analysis of earnings mobility is typically restricted to workers with earnings, excluding those without earnings, many of whom are women, children, or of retirement age. In contrast, it is typically assumed that each individual receives the (equivalized) total income of family (or household) to which he or she belongs. Because total household income is rarely zero, all individuals, regardless of age or labor market attachment, can in principle be included an analysis of income mobility. There is no universally correct definition of the income unit, and which should be used depends on the goals of the mobility analyst. For example, in a study of labor market flexibility, a focus on individual earnings is appropriate (though there remain questions about whether women can and should be included in such analysis—much empirical analysis is of men only). On the other hand, if the interest in mobility is stimulated by a desire to describe and summarize important features of society as a whole, then there is a strong case for using more inclusive samples. As we show later, some empirical studies focus on individuals of working age (variously defined), others on all individuals, and this can complicate cross-study comparisons.

"When" mobility issues refer to two aspects related to time. The first is the length of the period to which income refers to. For instance, is it the hour, week, month, or year? Economists often argue in favor of longer reference periods (e.g., a year) on the assumption that temporary variations and measurement error are smoothed out, thereby providing a more *accurate* measure of living standards. There is relatively little empirical evidence available about the veracity of this hypothesis because analysts rarely have income data for the same people over both shorter and longer periods. Böheim and Jenkins (2006) surveyed the literature and, from their analysis, argued that income mobility calculated using current (monthly) and annual income definitions are similar, and they provide a number of data-related reasons. Cantó et al.'s (2006) analysis is more comprehensive; based on comparisons from quarterly and annual income data for Spain, they show that use of the longer assessment period leads to higher estimates of poverty prevalence, lower inequality, and less mobility.

A second "When" issue relates specifically to mobility analysis in particular rather than income distribution analysis in general. For much mobility analysis, the data refer to a bivariate income distribution in which the marginal distributions refer to 2 years, t and $t+\tau$, and empirical analysis of longer-term inequality reduction requires a definition of how many years constitutes the longer term. In both cases, how far apart the base and final years are will affect the conclusions because the longer the interval, the greater the possibilities of mobility (as we illustrate later).[34] Choices about what interval to use have implications for the analysis that one can undertake too because data sets cover a time

[34] This issue is, of course, closely related to the issue of income reference period discussed earlier.

period of particular length (rarely more than 20 or 30 years), so researchers can only look at mobility trends if they use relatively short time windows for their measures. The constraint becomes acute with longitudinal data sets like EU-SILC (discussed later) in which the maximum time period is 4 years.

How researchers can address the three "W" issues is much constrained by the data that they have available to them, and this raises issues of comparability over time and country. Longitudinal data sources suitable for within-generation income mobility analysis are of two main types.

First, there are household panel surveys in which nationally representative samples of the private household population are interviewed about their incomes and many other domains of their lives in an initial year and then reinterviewed thereafter at regular intervals (usually a year). Second, there are administrative registers (e.g., tax files) in which income records for individuals are linked longitudinally. Household panel surveys typically utilize income definitions (i.e., resolve the "What" and to "Whom" issues) that are consistent with definitions accepted as being of good quality in large cross-sectional surveys. By contrast, administrative record data are typically designed for administration of the tax and benefit system, and the definitions used of income and the income-receiving unit, and the population that is represented, are determined by the needs of administration rather than by research. But register data also have advantages relative to surveys: Their samples are very much larger, issues of respondent dropout or measurement error do not arise in the same way (see the discussion later), and coverage of the very richest income groups is much better (they are typically not reached by surveys).

The clinching argument for empirical researchers in favor of household panel surveys over administrative registers is that the former became widely available for many countries, especially from the mid-1980s onward, with cross-nationally harmonized versions of the data following a few years later. Administrative registers with longitudinal income data have remained rare until recently in most countries, with the exception of Scandinavian countries, which have a rather longer history of use.

The longest-running household panel is the U.S. Panel Study of Income Dynamics (PSID), which began in 1968 and still continues, though it changed from annual interviewing to biennial interviewing after 1997. Panels started in the early 1980s in the Netherlands and Sweden, but the most well-known European panel is the German Socio-Economic Panel (SOEP), which started in the 1984 and is still running. Other country panels include the British Household Panel Survey (BHPS), which started in 1991 and finished in 2008. (The BHPS was recently replaced, after a break, by a new and very much larger panel (Understanding Society), which incorporates most of the original BHPS sample.) The Household, Income and Labour Dynamics in Australia (HILDA) survey began in 2001 and is ongoing. There is also Survey of Labour and Income Dynamics (SLID) for Canada, which is a rotating panel operational between 1998 and 2011.

As shall be seen later, it is the household panels cited in the last paragraph that have provided most of the empirical evidence about income mobility over the last two to three decades, both in their native format (often to examine trends over time within a country) or in a harmonized form (to undertake cross-national comparisons). The production of cross-nationally comparable household panel data with harmonized labor earnings and household income variables has been one of the major successes in social research infrastructure creation over the last few decades.

The Cross-National Equivalent File (CNEF) began in 1991 with harmonization of data from the U.S. PSID and German SOEP and incorporated the BHPS and SLID in 1999 and HILDA in 2007. (Data for more countries have been added subsequently.) It should be stressed that the project does more than simply harmonize variables; it adds value. One important example of this is the derivation of comparable posttax posttransfer household income variables. The original PSID family income variable refers only to pretax posttransfer income, and the government transfers do not include income derived from nonrefundable tax credits (the EITC) or near-cash benefit income in the form of Food Stamps (now called SNAP). The CNEF uses the NBER TAXSIM model to simulate taxes. Similarly, involvement in the CNEF project was a stimulus for the SOEP to develop and maintain a similar model in-house. (Other CNEF members also use such models.) For a more detailed discussion of the CNEF, see Frick et al. (2007).[35]

Another important initiative providing cross-nationally comparable panel data on incomes was the former European Community Household Panel (ECHP), though this has been used less often for mobility analysis than the CNEF and its constituent panels. The ECHP relied on "input" harmonization by contrast to the CNEF's "output" harmonization. That is, household panel surveys with the same design and questionnaires including the same variables were fielded in a number of countries, so that harmonization was built in from the start. Data from a maximum of eight annual interview rounds are available, covering the period 1994–2001. Twelve EU member states participated in the ECHP initially, with two more joining shortly thereafter. The ECHP never realized its full potential because, for many years, researcher access to the data was constrained and financially costly. This is in contrast with the CNEF, which from the start has had a much more open data access policy and has been more research(er) driven.[36]

The ECHP was replaced—after a gap—by the European Statistics on Income and Living Conditions (EU-SILC) from 2005. EU-SILC is explicitly designed to deliver data on a set of social indicators that include income distribution statistics. This is output

[35] For documentation and user access information, see http://cnef.ehe.osu.edu/ or http://www.human.cornell.edu/pam/research/centers-programs/german-panel/cnef.cfm.

[36] The Survey of Health, Ageing, and Retirement in Europe (SHARE) is another multicountry longitudinal study proving input-harmonized income data, and also research driven. Its focus, however, is on older individuals, and so it cannot be used to study income mobility in the wider population.

harmonization again, though the target variables are predefined by the needs of EU policy making rather than by researchers. Some member states use administrative registers to produce the data; others use panel surveys, an aspect that has led to questions about data comparability (see later). The longitudinal data in the publicly released EU-SILC data sets track individuals for a maximum of 4 years (by design), and so the scope for longer-run mobility analysis is ruled out. The great advantage of the EU-SILC longitudinal data is that, when mature, they will cover all EU member states. Understandably the EU-SILC has not been much used for income mobility to date, and this is reflected in our review of evidence that follows.

This review of data sources suggests that there has been a substantial increase over the last three decades in the volume of high-quality longitudinal data available to researchers. But there remain a number of important issues of empirical implementation that need to be kept in mind when assessing the value of a particular mobility study. So, before turning to discuss empirical evidence, we briefly review these issues.

There are generic issues associated with longitudinal surveys, notably the potential problem of survey attrition. Over time, some respondents to a panel survey drop out from the data, either no longer wishing to participate or unable to be tracked down for interviews. Attrition has two potentially adverse effects. The first is reduction of sample size, with consequences for the precision of estimates. The second potential effect, more commonly discussed, is on the representativeness of the sample. Particular groups such as young people tend to be more likely to drop out, in which case estimates may be biased.[37] Differential attrition may be related to both observed and unobserved characteristics of individuals and families. For the former case, data producers routinely produce and release sets of weights that can be used to maintain the representativeness of estimates, and virtually all the studies cited in our evidence review use these weights. By definition, it is harder to assess the effects on estimates of differential dropout related to unobserved characteristics; it requires modeling of the attrition process. For an extensive discussion of attrition in U.S. household panel surveys, see Fitzgerald et al. (1998) and other papers in the Summer 1998 issue of the *Journal of Human Resources*.

The likely impact of attrition is associated with the type of mobility analysis undertaken. Attrition between successive waves of a household panel is typically relatively low (around 5%) with the exception that dropout rates are noticeably greater between the initial and second waves. Estimates of mobility over short periods (1 or 2 years, say) are likely to be less affected by attrition than estimates based on long runs of data.

[37] Representativeness typically refers to the ability of the sample to represent the private household population in the first wave of the panel. If a country experiences significant migration or immigration, a panel inevitably becomes unrepresentative of the population in later years. Sample refreshment has been employed to counter this problem, but if mobility estimates are required for time points spanning the old and new population structures, refreshment cannot improve representativeness.

Respondents may remain in a longitudinal survey, but not provide complete responses to particular questions, either because they do not understand the question, do not know, or do not wish to provide the answer. This is the issue of "item" nonresponse leading to missing data for some respondents and, as with attrition, may be associated with both observed and unobserved respondent characteristics. Item nonresponse is particularly prevalent for questions about income sources by comparison with items such as, e.g., a respondent's age. In the public-use panel data sets used by mobility researchers, missing income values are typically replaced by an imputed value (together with a flag that enables identification of such observations) generated using procedures allocating similar values to respondents with similar sets of (observed) characteristics. Imputation is very useful for analysts but can potentially have effects on analysis because, by comparison with nonimputed data, extra "noise" is added by the inevitable imperfection of the process.[38] These can have particular effects on mobility analysis because some of the changes in a person's income over time may simply reflect the imputation process in the different years. But if one simply drops the imputed observations, there may be a critical loss of sample size and use of a potentially nonrepresentative subsample. In most of the income mobility studies discussed later, analysts have routinely used imputed data on household income. By contrast, in studies of earnings volatility, it is a more common practice to drop imputed observations. Researchers tend to find that this reduces observed volatility, but the effects are relatively small. Again, the likely effects will depend on whether the particular mobility measure employed requires, say, two years relatively close together, or many years over a longer interval.

The problems raised by imputation are closely related to the more general issue of measurement error in earnings and income data. Even if survey participants respond to a question, their answer may be incorrect either because the respondent does not want to give the true answer or simply does not know what it is. Key issues are whether observed responses are systematically under- or overreports of the (unobserved) true value or simply random, and how errors are correlated across successive years of data for the same respondent. Clearly, the answers to these questions may differ by income source. The largest body of research on measurement error has been about labor earnings and used validation studies in which linked administrative record data are used to provide a picture of each worker's "true" earnings (see, e.g., the survey by Bound et al., 2001). Few studies have looked at the effects of measurement error on measures of earnings mobility.

The perhaps surprising finding of Gottschalk and Huynh (2010) is that estimates of men's earnings mobility, defined in terms of the Pearson correlation between log earnings in 1 year and the next, are much the same in the survey data and their administrative data

[38] The imputation of households' tax payments when deriving measures of posttax posttransfer income is another important example of useful imputation that may also add noise.

set. The result arises not because measurement errors are not important; rather, it is because they are "nonclassical" in nature (i.e., mean-reverting and correlated across years), and these various features happen to offset each other. See also Fields et al. (2003), who use a nonclassical measurement error model similar to that of Gottschalk and Huynh (2010) to put bounds on estimates of income change. For the case considered, they argued that the effects of measurement error are "relatively minor" (Fields et al., 2003). Dragoset and Fields (2006) calculated a large portfolio of mobility measures from both survey and linked administrative record data on U.S. men's earnings. They concluded that most of their qualitative results are the same in both data sources, and that the estimates from the administrative source were neither systematically above nor below the corresponding survey estimates. Overall, this small body of research might be taken to imply that measurement error has relatively unimportant effects on measures of mobility in practice. We would caution against this interpretation, convenient as it is for empirical researchers; the situation is more that we know rather little at present. All the studies cited refer to earnings for U.S. men, and results may differ for household income and in other countries. (The only similar study for household income that we are aware of is by Rendtel et al, 2004, who also reported finding mean reversion and serial correlation.) There is also a more fundamental question of whether administrative record data can be assumed to provide error-free representations of the truth (Abowd and Stinson, 2013).[39]

A rather different sort of measurement error arises in the case of outlier observations, for example, very high or very low observations. These may be genuine but may also represent errors of, e.g., transcription leading to additional zeros being added. The problem is that even if the number of observations with this kind of data is very small, they may have a big influence on the estimates that are derived. This lack of robustness is undesirable. See Cowell and Schluter (1999) for a discussion of this problem in the context of income mobility analysis. Empirical analysts' response to this issue is usually to simply drop a fraction (e.g., 1%) of the very richest and of the very poorest income values in each year. This procedure, known as "trimming," or similar algorithms directed at removing potential outliers, has been applied in virtually every study cited in our discussion of empirical evidence.

A final empirical issue is whether income changes over time represent genuine mobility or, instead, systematic changes associated with life cycle patterns of such earnings following an inverse-U shape with age. Many income mobility studies do not adjust for this

[39] Aside from analysis based on validation studies, there have been a small number of model-based assessments of the impact of measurement error on estimates of poverty transition rates: see, e.g., Breen and Moisio (2004) and references therein. Longitudinal data on observed transitions are combined with an assumption that the "true" transition probabilities are stable over time, so that difference between them is attributed to measurement error. In technical terms, the statistical approach involves fitting latent class models with a Markov structure. For further discussion, see Jenkins (2011a, pp. 53–55).

factor; they look at observed incomes. Some other studies, mostly of earnings mobility, have regressed observed earnings against variables such as age, and then the mobility analysis is of the earnings residuals: see later discussion.

10.4.2 Intragenerational Income Mobility in the United States: Levels and Trends

We take as our initial reference point the estimates of income mobility for the United States provided by Hungerford (2011), as he uses good-quality comparable data from PSID (as released via the CNEF) and provides a range of mobility summaries. (Transition matrices from the study were presented in Table 10.1 earlier.) Hungerford compared mobility over two 10-year intervals, 1979–1988 ("1980s") and 1989–1998 ("1990s"). The measure of income is annual disposable (posttax posttransfer) family income adjusted for differences across families in household size and composition using the equivalence scale proposed by Citro and Michael (1995). His samples include all individuals within households. In the 1980s sample, about half the SEO sample was dropped in 1997. All estimates are derived using the PSID's weights. We noted earlier that, in both periods, there appeared to be substantial short-distance mobility over a 10-year period, but long-distance moves were relatively rare. Moreover, the chances of upward mobility from the bottom and downward mobility from the top appeared symmetric. We now compare mobility in the two decades in greater detail, in particular considering whether mobility increased or decreased according to various mobility concepts and measures.

To assess changes in positional mobility, a natural first approach is to apply the dominance check of Atkinson and Bourguignon (1982) based on the differences in the discrete cumulative densities implied by the decile transition matrices in Table 10.1. See Table 10.2 for the estimated differences. First-order dominance does not hold; there is a mixture of positive and negative differences.[40] There is an interesting pattern, however. Most of the positive differences (greater cumulative density in the 1980s) are found in cells corresponding to movements out of or into the poorest fifth of the distribution. Put another way, there is greater movement in the 1980s than the 1990s into and out of the richest 80%, broadly speaking.

Saying conclusively that mobility increased or decreased in the United States between the 1980s and 1990s, and by how much, requires additional assumptions about the weighting of mobility in different parts of the distribution. Also, the answers depend on the mobility concept. These points are illustrated by the mobility index estimates reported by Hungerford (2011) and summarized in Table 10.3. The first three rows

[40] The density estimates and conclusions drawn from them need to be interpreted cautiously, not the least because they are susceptible to measurement error and sampling variability. If the estimates in Table 10.2 are rounded to 2 d.p. to reflect this (rather than 3 d.p. as reported), then many matrix entries become zero, and there is now dominance; positional mobility is greater in the 1980s than the 1990s.

Table 10.2 Differences in cumulative density: United States, 1979–1988 versus 1989–1998

Origin group	Destination group 1	2	3	4	5	6	7	8	9	10
1	0.2	−0.1	−0.2	0.0	0.3	0.2	0.1	0.1	−0.1	0.0
2	0.0	0.0	0.4	0.6	0.5	0.2	0.2	0.1	0.0	0.0
3	−0.2	−0.5	−0.2	0.0	0.0	−0.5	−0.1	−0.1	0.0	0.0
4	−0.2	−0.7	−0.6	−0.6	−0.7	−0.7	−0.2	−0.3	0.1	0.0
5	0.0	−0.3	−0.3	−0.5	−0.7	−0.5	0.0	−0.1	0.4	0.0
6	0.1	−0.1	−0.1	−0.4	−1.1	−1.3	−0.9	−0.5	0.4	0.0
7	0.1	0.2	0.0	−0.3	−0.8	−0.9	−0.8	−0.3	0.3	0.0
8	0.1	0.2	−0.2	−0.2	−0.3	−0.7	−1.1	−0.7	−0.3	0.0
9	0.0	−0.1	−0.3	−0.2	−0.4	−0.4	−0.7	−0.6	−0.6	0.0
10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Note: The estimates are in percent, rounded to one decimal place, and show in each cell the cumulative discrete density for the 1980s minus the corresponding cumulative discrete density for the 1990s.
Source: Authors' calculations from Hungerford (2011, Tables 2 and 3), based on PSID data.

Table 10.3 Selected mobility indices (%): United States, 1979–1988 versus 1989–1998

Index	1979–1988	1989–1998
Decile mobility	79.1	77.0
Normalized trace	87.9	85.6
Gini mobility	36.2	34.4
Equalization (Shorrocks, Gini-based)	10.9	11.1
Equalization (Fields, Gini-based)	2.1	8.2
Average of absolute income changes (*D*l)	11,368	13,878
Average of absolute income share changes	0.421	0.459

Note: The estimates are in percent, rounded to one decimal place, apart from those in the last two rows (in constant-price dollars). Decile mobility is the proportion of persons changing at least one decile group. The normalized trace is the Shorrocks (1978b) index calculated from the decile transition matrix. The Gini mobility index is the index of Yitzhaki and Wodon (2005). The Equalization indices are those of Shorrocks (1978a) and Fields (2010). On the average of absolute income and income share changes, see Fields and Ok (1996) and Fields (2010). See text for more details.
Source: Authors' calculations from Hungerford (2011, Tables 4 and 8, and p. 97), based on PSID data.

of the table provide estimates of positional mobility (reranking), and all the indices show a small decline between the 1980s and 1990s. In contrast, the Shorrocks and Fields equalization indices record an increase, and so too do the two measures of income flux shown in the bottom two rows. For the last four indices, the estimated increase is small, with the exception of the Fields equalization measure, for which the large change reflects the increase in (cross-sectional) income inequality over the period. The general lesson is that conclusions about whether mobility increased or decreased between the 1980s and 1990s depend on the mobility index employed.

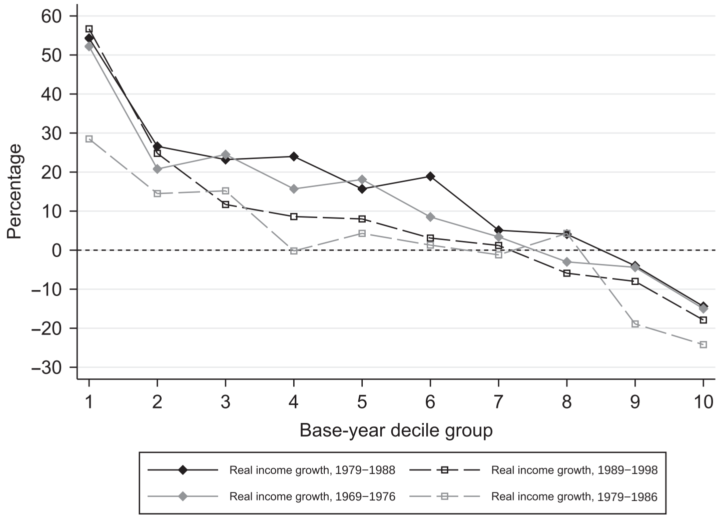


Figure 10.9 Median real income growth, by base-year decile group: United States, by period. *Note: The estimates show median income growth for each base-year decile group over the relevant period. Source: Hungerford (1993, Table 9) and Hungerford (2011, Tables 5 and 6).*

Mobility as individual income growth is also summarized by Figure 10.9, which shows the median real income growth for each base-year decile group, by period. (This is a grouped data version of Figure 10.7 discussed in the previous section.) Clearly income growth is pro-poor in the United States (consistent with regression to the mean), but the patterns differ between the 1980s and 1990s. Income growth was greater in the 1990s than the 1980s for the richest eight base-year decile groups, but no different for the two poorest base-year decile groups.

The extent to which U.S. mobility comparisons can be extended to periods before the 1980s and after the 1990s is restricted by data availability (e.g., the PSID only started in 1968) because different studies use different income variables and estimation samples and often do not report the same mobility statistics.

For example, Hungerford (1993) provides much information about U.S. income mobility in the 1970s and 1980s, but the estimates are not fully comparable with those in Hungerford (2011) because the earlier study uses a different income definition (pretax posttransfer income rather than equivalized posttax posttransfer income), and the interval between base- and final-years differs (8 years rather than 10; e.g., 1979–1986 rather than 1979–1988). The relevance of definitional differences is illustrated by the estimates for the "1980s" from the two studies of the proportions of individuals remaining in the poorest 10th and remaining in the richest 10th: 44.3% and 40.0% according to Hungerford (2011), but 49.0% and 42.1% according to Hungerford (1993, Tables 1 and 2). Look also at the different estimates of real income growth rates for the 1980s for the two periods in

Figure 10.9. Using the Hungerford (1993) definitions, the overall growth rate for the 1980s is smaller (which is unsurprising because aggregate income growth was positive throughout the mid-1980s; Hungerford, 2011, Table 1), but observe that the estimates of pro-poorness in income growth also differ (the income growth curves from the two studies do not have the same slope).

One can compare mobility in the 1970s and the 1980s, however. If we examine differences in cumulative densities using Hungerford's (1993) estimates, again there is no clear-cut mobility ordering (authors' calculations), and there is a broadly similar pattern of differences to that described earlier. Hungerford (1993) does not report summary indices to compare with those in Table 3, but two statistics based on the transition matrices (Cramér's V) and the contingency coefficient "are the same . . . suggesting that the degree of association between a person's decile rank in one year and another was the same in the 1970s and 1980s" (Hungerford, 1993, p. 407). Fields and Ok (1999a) used exactly the same data as Hungerford (1993) and reported that their measure of income flux, the average of the absolute changes in log income, increased from 0.498 in the 1970s to 0.528 in the 1980s.[41] So, again, changing the mobility concept leads to a different conclusion about trends.

Hungerford's (1993) study is also useful because it analyzes whether the estimated mobility patterns are robust to adjustment for transitory income variation. Specifically, Hungerford calculated each individual's 5-year longitudinally averaged income (centered on the year in question) and used these "permanent" incomes instead of the single-year incomes to define base-year and final-year income positions. Interestingly, the patterns of mobility revealed are remarkably similar, though with perhaps less movement at the top and bottom of the distribution.[42] For example, according the annual income calculations for 1979–1986, 12.9% of the poorest fifth remain in that group and 11.0% of the richest fifth remain in that group. According to the permanent income calculations, the corresponding estimates are 11.5% and 9.6% (authors' calculations from Hungerford, 1993, Tables 2 and 4).

To examine trends in U.S. income mobility further, we turn to Bradbury (2011). She provides estimates using consistent definitions for the period 1969–2006 and for a large portfolio of mobility indices. Her estimates are not fully comparable with Hungerford's, however. Although she and Hungerford (2011) both use posttax posttransfer real family income measures from the CNEF version of the PSID, they use different samples. Bradbury focuses on adults who are a family head or spouse rather than all individuals

[41] Fields and Ok's (1999a) decompositions reveal that the increase in income movement is entirely accounted for by persons with education to high school level or above and by young adults rather than prime-age adults.

[42] One potential noncomparability is that the estimation samples differ slightly; the permanent income estimates are based on balanced samples with valid data for all 5 years within the relevant period.

within families, both head and spouse (if present) are required to be of working age (16–62 years), and the time interval spans 11 years rather than 10. She used the square-root-of-household-size equivalence scale rather than the Citro and Michael (1995) one.

Trends in three general indices of positional mobility are displayed in Figure 10.10: the fraction of individuals changing decile group ("decile mobility"), one minus Spearman's rank correlation, and Yitzhaki and Wodon's (2005) Gini mobility index. All three indices are broadly constant over the 1970s and decline over the 1980s (11-year intervals starting at the end of the 1970s), with the rate of decline perhaps slowing from the late 1980s onward. The fall in mobility over the 1980s is consistent with Hungerford's estimates of trends based on only two intervals during this period, but is rather larger in magnitude. The Gini mobility index fell by about a sixth between the intervals starting in 1979 and 1989 (but only about 5% according to Hungerford, 2011). One minus the rank correlation fell by about one-fifth over the same period, and so the decline in positional mobility is relatively large. It is unclear what lies behind the secular decline in mobility, but we note that it was at the end of the 1970s that U.S. family income inequality also began to increase (Burkhauser et al., 2011), suggesting that inequality and positional mobility share some common drivers. There is no very obvious association between series' turning points and the business cycle (there were recessions at the beginning of the 1970s and 1980s).

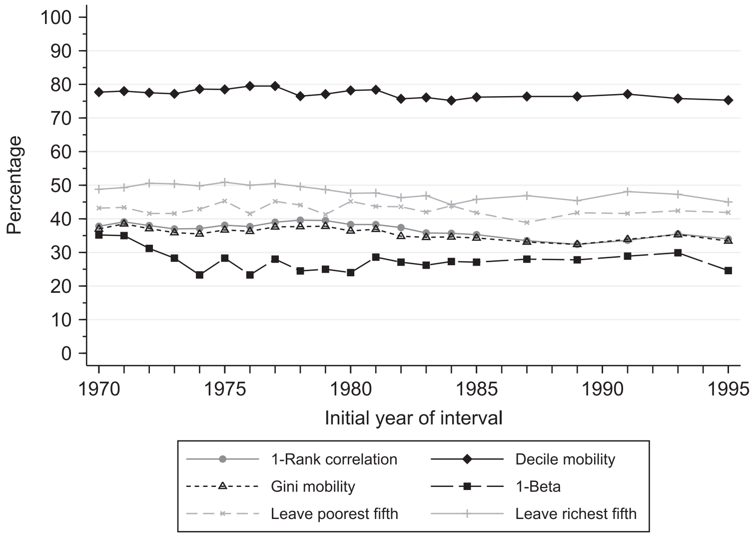


Figure 10.10 Indices of positional income mobility: United States, 1970–1995. *Note: The estimates refer to 11-year intervals, with incomes in base- and final-year averaged over 2 years. For example, the estimates labeled as 1970 refer to incomes longitudinally averaged over 1969 and 1970 (base year) and 1979 and 1980 (final year). See text for index definitions. Source: Bradbury (2011, Tables 2 and 3).*

The conclusions about trends cited so far refer to income changes over an interval of 10 or 11 years, and it is of interest to know how results change if rather different interval lengths are used. The research of Gittleman and Joyce (1999) suggests some sensitivity. Using PSID data for 1967–1991 and, like Bradbury (2011), focusing on working-age adults and employing a broadly similar income definition,[43] they calculated IRs, defined as the percentage of individuals remaining in the same fifth, for intervals of 1, 5, and 10 years. Gittleman and Joyce (1999, Table 1, Figure 2) show that the level of positional mobility increases (the IR falls) as the interval width is widened. But conclusions about mobility trends are also affected. For the 10-year interval case, there is a small downward trend during the 1980s consistent with Bradbury's (2011) estimates. However, 5-year IRs exhibit no similar trend, and 1-year IRs generally decline from the end of the 1960s until the end of the 1970s and increase in the following decade (though the changes are not large in absolute magnitude).

To provide a comparison with another commonly used mobility index, we also show trends in one minus Beta. It follows a different trend, which is perhaps unsurprising given that it is not a purely positional measure (see Section 10.3). Compared to the trends shown by the three positional indices, the decline during the 1970s is earlier and sharper, and there is no decline during the 1980s.

The final two measures shown in Figure 10.10 are two "corner probabilities" from a quintile transition matrix (cf. Section 10.3), specifically the proportion of individuals in the poorest fifth in the base-year who are in a different fifth in the final year and, analogously, the proportion leaving the richest fifth over the relevant interval. These statistics pick up on particular aspects of positional mobility. Interestingly, it appears that the trend in the percentage leaving the richest fifth tracks the trend in overall positional mobility better than does the trend in the proportion leaving the poorest fifth. The estimates also bear on our earlier comments that the U.S. decile transition matrices for the 1980s and 1990s suggest that there is a broad symmetry to upward and downward mobility. We now see that asymmetry is more apparent if mobility is summarized using quintile rather than decile groups. In particular, it appears from Bradbury's (2011) estimates that the chances of downward movement from the top (richest fifth) are typically several percentage points greater than the chances of upward mobility from the bottom.

This asymmetry finding also may be contingent on the particular samples and other definitions used. For example, Bradbury and Katz (2002, Annex A) report quintile transition matrices for 1969–1979, 1979–1989, and 1988–1998 using similar PSID samples to Bradbury (2011), except that "working age" now refers to a wider age range (head and spouse [if present] less than 66 years), and family income is pretax postgovernment family income, equivalized using the PSID scale. The two probabilities are approximately equal

[43] But see later for more about differences.

in each matrix (50% in the first two periods, 47% in the last one). In contrast, Gittleman and Joyce (1999, Table 5) report quintile transition matrices for 1967–1979 and 1979–1991 using a similar income definition (but equivalized using the U.S. poverty line), and "working age" refers to head and spouse between 25 and 65 years. According to this study, the chances of leaving the poorest fifth are distinctly smaller than the chances of leaving the richest fifth (around 50% compared to around 60%).

Trends in mobility defined as equalization of longer-term incomes are summarized by Figure 10.11 using Shorrocks's (1978b) measure $M = 1 - R$. The long series (shown in black) are derived from Bradbury (2011); we discuss the series in gray shortly. Although mobility levels differ substantially depending on which inequality index is used—there is much greater mobility according to the Theil index compared to the Gini—the patterns of change over time are the same according to the two series. There was a decline in mobility between the early 1970s and the mid-1980s, followed by a rise over the following decade, with leveling off around the mid-1990s. Although the changes are small in absolute terms, they are relatively large in proportionate terms. For example, between the mid-1980s and mid-1990s, the Theil-based measure increased by some 15% and the

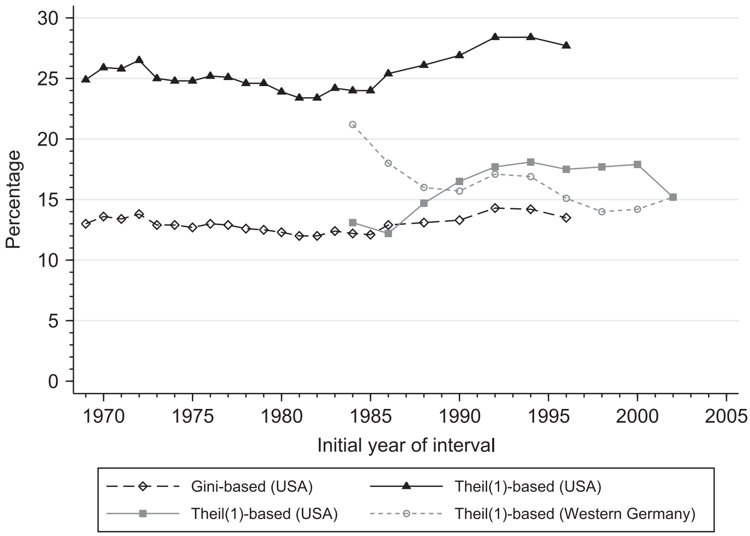


Figure 10.11 Mobility as longer-term income inequality reduction: United States, 1970–1995. *Note: The estimates refer to the Shorrocks equalization measure, $M = 1 - R$, calculated using the Gini and Theil inequality indices. The Bradbury (2011) calculations are based on 11-year intervals with longer-term average incomes calculated using every second year's income to handle the PSID's change to alternate-year interviewing in the late-1990s. The Bayaz-Ozturk et al. (2014) calculations use 5-year intervals, with interval base-years 2 years apart. Sources: Bradbury (2011, Table 4) for the series shown in black and Bayaz-Ozturk et al. (2014, Table A1) for the series shown in gray. Both use PSID (CNEF) data.*

Gini-based measure by almost 13%. The results are consistent with Hungerford's (2011) finding of only a small increase in a Gini-based measure between the 1980s and 1990s, but Figure 10.11 shows that this is partly a consequence of the timing of measurement; Hungerford's two intervals lie on either side of the bottom of a U-shaped series. Also, the turning points in the these two series differ from those for the positional measures shown in Figure 10.10, suggesting that the different aspects of mobility have different underlying causes. In addition, mobility according to the Shorrocks measure is much the same (Gini-based index) or greater (Theil-based index) in the mid-1990s than in the early 1970s, whereas mobility is lower according to the positional mobility indices shown in Figure 10.10.

The research of Bayaz-Ozturk et al. (2014) allows to consider what happened to mobility as equalization after the mid-1990s. Although they also use a Theil-based measure, similar income measures, and the same data source, their series are not directly comparable with all individuals in families in their analysis samples (not only working-age adults) and they use a 5-year rather than 11-year interval. As a consequence, mobility levels are estimated to be substantially lower in all years (compare the gray line for the United States with the black one). Reassuringly, however, the series show broadly similar trends (and turning points) over the period for which they overlap. Bayaz-Ozturk et al.'s (2014) estimates indicate that mobility changed little in the second half of the 1990s, with a suggestion that it fell again in the 2002–2006 period.

All estimates of trends in household income mobility presented so far in this section are based on PSID data, and it is of interest to know whether the evidence from other data sources tells a similar story. The main reference point on this issue is Auten and Gee's (2009) work based on income data from tax administration records covering the two decades between 1987 and 2005. The data and definitions used are not fully comparable with those in the PSID studies, but there are advantages from having much larger sample sizes and much better coverage of top incomes. The analysis focuses on tax filers and their spouses (if present), excluding taxpayers aged under 25 years. An individual's income is the income of the his/her tax filing unit, divided by the square root of household size. Income is a measure of pretax income and includes all taxable income sources reported on tax returns supplemented with data about Social Security benefit income provided to the Internal Revenue Service.

The first part of Auten and Gee's (2009) article describes mobility between 1996 and 2005 in terms of positional mobility (transition proportions) and income growth (by base-year income group). The results are broadly consistent with the studies cited earlier in terms of pointing to substantial movement between quintile groups but with short-distance moves the most prevalent, and real income growth is greater the poorer the base-year income group. The distinctive feature of the study is the information about mobility at the very top of the distribution with mobility statistics also provided for the very top income groups. The authors report that there is a large amount of turnover at the top and that "the incomes of many taxpayers at the highest levels are very volatile" (Auten and Gee, 2009, p. 311). For example, among the richest 0.01% in 1996, only 23% remained in the group in 2005. Although over 80% were still in the top 1%, 6% dropped out of the richest fifth (Auten and Gee, 2009, p. 311).

The second part of Auten and Gee's (2009) article assesses changes in mobility between 1987 and 1996 and 1996 and 2005 using the same measures, and the authors state with regard to positional mobility that "the basic finding . . . is that [it] is approximately the same in the last 10 years as it was in the previous decade" (Auten and Gee, 2009, p. 311). Also, although overall real income growth was around 23% in the first decade compared to 8% in the second, its pro-poor pattern was similar across most of the distribution. (Median real income increased by about 15% points for the top four quintile groups and by about 10% points for the poorest base-year fifth). Things were different at the very top, however. Real income growth was −32% for the top 1% in 1987, and −31% for top the 1% in 1987 (Auten and Gee, 2009, Table 10.7).

Further information about persistence in the top 1% is provided by Auten et al. (2013) for tax filers aged 25–60. Their Table 10.3 shows survival rates in the top 1% (i.e., taking taxpayers in this group in some base year t, what proportion of them were in the top 1% in each and every subsequent year $t+\tau$, where $\tau=1, 2, 3, 4, 5$). Base years run from 1991 to 2009. The 5-year survival rates range between 21% and 36% and the 1-year survival rates between 52% and 70%. The authors pointed out that lower persistence rates tend to occur in recessionary periods (1991, 1999 through 2001, and 2007), and they suggest that income sources of particular relevance for the richest groups such as capital gains and net business income are relatively sensitive to the business cycle.

The body of evidence on trends in measures of mobility as family income risk is much smaller than for the other concepts and also is difficult to synthesize because a wide range of descriptive and model-based measures has been used. One set of PSID-based estimates derived by Gottschalk and Moffitt (2009) is shown in Figure 10.12. The estimates refer to all individuals in families, and income is the PSID pretax posttransfer measure equivalized using the U.S. poverty line for the family type in question. The chart shows that the transitory variance of log annual family income increased substantially, by around 70%, between the mid-1970s and 2000, though this included a period during the 1980s when there was little change. Other PSID-based studies report a similar rise taking the period as a whole (and concur on the increase during the 1990s), though they use different measures, time periods, and analysis samples. See inter alia Hacker and Jacobs (2008) and especially Dynan et al. (2012), who also include a useful review of earlier studies for the United States.

There is ongoing debate about the robustness of the PSID-based estimates, notably for the 1990s onward. This is illustrated by the findings of Dahl et al. (2011). They assess household income volatility using data in which responses to the Survey of Program Participation are linked to earnings data from Social Security Administration records

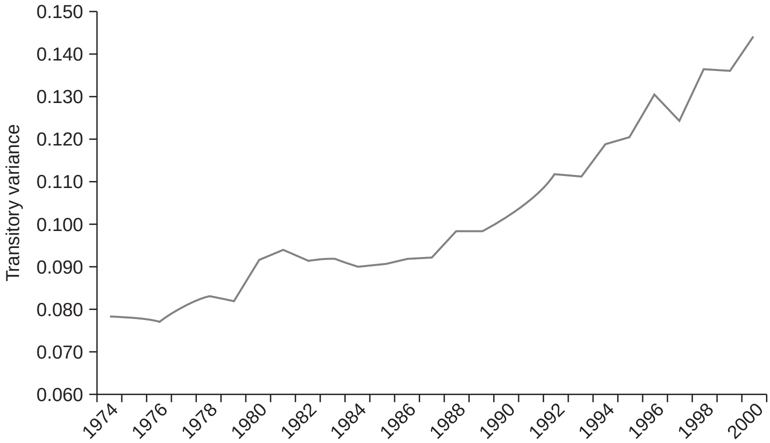


Figure 10.12 Transitory variance of log annual family income: United States, 1974–2000. *Note: Transitory variances computed using the Gottschalk and Moffitt (1994) window-averaging method, with rolling 9-year windows. Source: Gottschalk and Moffitt (2009, Figure 5), based on PSID data.*

("SIPP-SSA" data).[44] Household income is calculated as the sum across household members of earnings from the SSA records plus the survey reports of nonlabor income (but income is apparently not equivalized), and the analysis samples refer to individuals in households with heads aged 25–55 years. Using multiple SIPP panels, the authors derived 1-year volatility estimates at eight time points between 1985 and 2005. The headline finding is that there is no upward trend in volatility, and in particular there is little change over the 1990s. Dahl et al. (2011, p. 769) conclude that they cannot reconcile their results with the divergent set of results from the PSID and other survey data sources, but draw attention to the potential roles of differences in the data per se (rather than the summary measures applied to them). Reconciliation of results is an important task for future research.

The recent study of DeBacker et al. (2013) is a helpful contribution in this respect. It is based on a 1/5000 sample of the U.S. taxpayer population with panel data covering 1987–2009, analyzing individuals aged 25–60 years. There are no potential issues arising from matching or imputation for missing values as in the SIPP-SSA data. The definition of household income is similar to the Auten and Gee (2009) one (see earlier). The authors calculate 1- and 2-year volatility measures and the transitory variances using descriptive and model-based estimates. According to all three measures, there was a small rise throughout the period considered (Figures VI, VII, A.1(e)). DeBacker et al. (2013)

[44] Between 10% and 20% of respondents were not matched with SSA records and up to 40% in the 2001 SIPP panel (Dahl et al., 2011, p. 755). This is a potential source of bias and one that the authors were unable to address.

attributed the rise in the transitory variance primarily to changes in spousal labor earnings and investment income.[45]

We finish this discussion of U.S. mobility trends with reference to evidence about the mobility of individual labor earnings. The recent literature on trends is dominated by analysis of what we have described as measures of income "risk," as summarized by the transitory variance and volatility of earnings, and is almost entirely about men's earnings. (The estimates for household income risk cited earlier are usually by-products of this analysis.) Most analysis is of earnings residuals rather than raw earnings. That is, researchers first run regressions to control for differences in education, age, and work experience and work with the residuals from the fitted models.

Most studies show that men's earnings instability increased during the 1970s, but then leveled off somewhat through to the early- to mid-1980s or fell slightly. Findings about what happened in the 1990s and 2000s depend on the data set and measure used. This is particularly so when measures of volatility are used. Estimates derived from the PSID suggest a rise in volatility (Celik et al. (2012), Shin and Solon (2011), Moffitt and Gottschalk (2012)), whereas those derived using linked-CPS data, administrative record data, or survey data linked to administrative record data, suggest that volatility either remained flat (Celik et al., 2012; Dahl et al., 2011; DeBacker et al., 2013; Ziliak et al., 2011) or at least appear not to have risen (Juhn and McCue, 2010). There appears to be more agreement across studies and data sets about what happened in the 1990s and afterward if the focus is on the transitory variance of men's earnings rather than volatility, namely that the earlier rise leveled off in the 1990s and thereafter: see, e.g., Gottschalk and Moffitt (2009), Moffitt and Gottschalk (2012), and DeBacker et al. (2013). This is consistent with a finding that it is the variance of the permanent component of men's earnings that has grown most over this period, and note that measures of short-term volatility reflect permanent as well as transitory shocks. For further discussion of the different findings across measures and data sets, see Moffitt and Gottschalk (2012, Section V).

For analysis of trends in earnings mobility using other measures, we refer to Buchinsky and Hunt (1999) and Kopczuk et al. (2010). (There are few other relatively recent studies.) Buchinsky and Hunt (1999) is a detailed study of mobility of in wages and annual labor earnings over the period 1981–1991 using the cohort of young people in the National Longitudinal Study of Youth (aged 14–24 in 1979), excluding military personnel and individuals who are self-employed or in education. Mobility is summarized using the Shorrocks equalization measure (M, using multiple inequality indices) and transition probabilities estimated using the nonparametric density method cited in Section 10.3. The main result about trends is that mobility declined between 1981 and 1991, regardless of which inequality index M is calculated with and using window lengths of 1, 4, or 6

[45] Although the transitory variance increased, DeBacker et al. (2013) emphasized that it was the increase in the permanent variance that contributed most to the increase in inequality over the period.

years (Buchinsky and Hunt, 1999, Table 2). Positional mobility also declined; the chances of remaining in the same quintile group, and the average jump and normalized trace indices also fell. The decline in mobility as equalization is the opposite trend from what we discussed earlier for household income. One potential reason relates to the fact that this is a youth cohort, and Buchinsky and Hunt (1999) discussed the difficulties of separately identifying time and age effects.

Kopczuk et al. (2010) is a landmark study of earnings mobility because of its rich data. They used longitudinal Social Security Administration data on earnings stretching from 2004 right back to 1937. The focus is on men and women aged 25–60 years with annual earnings from employment in the commerce and industry sectors greater than a minimum threshold (one-fourth of the full-time full-year minimum wage in 2004 indexed forward and back).[46] Kopczuk et al. (2010) exploited their long series to examine trends in mobility with multiple short- and long-term measures. They variously used longitudinally averaged earnings over 5- and 11-year windows and looked at measures defined for intervals of various length between base and final year. With their large samples and coverage of the tax data, they could also analyze mobility at the top of the earnings distribution.

Short-term mobility is summarized using three measures, the rank correlation for earnings 1 year apart, and a Gini-based Shorrocks rigidity measure ($R = 1 - M$) and transitory variance of log earnings (calculated using a method similar to the BPEA one), each derived using income averaging over moving 5-year windows. According to the first two measures (Kopczuk et al., 2010, Figures IV, V), earnings mobility for all workers increased sharply over the years of World War II and then fell, reaching prewar levels by around 1960. Thereafter, there was remarkably little change. The transitory variance for log earnings was also roughly constant from around 1960 until the mid-2000s. This result is at odds with the PSID estimates for 1970s discussed earlier (see, e.g., the increase shown by Gottschalk and Moffitt, 2009, Figure 1), but consistent with the IRS-data based study from 1987 onward by DeBacker et al. (2013) (which also, like Kopczuk et al.'s 2010 study, emphasizes the increase in the permanent rather than transitory variance).

From 1978 onward when earnings data were no longer top-coded, Kopczuk et al. (2010, Figure 6) examined the probabilities of remaining in the top 1% over one-, three-, and 5-year intervals. There is remarkable stability in these series (e.g., the 1-year probability ranges between 72% and 79%, and the 5-year probabilities between 60% and 65%). These staying probabilities are greater than those shown by Auten et al. (2013, Table 3) for pretax income (for 1991–2009). It is the pretax income components other than labor

[46] The authors undertake extensive checks of the sensitivity of their findings to different assumptions about sample selection, top-coding, coverage of the various administrative sources, and so on and report that their conclusions are robust.

income that are apparently sensitive to the business cycle (and note also that Kopczuk et al.'s (2010) series predate the onset of the Great Recession in 2007/2008).[47]

To summarize long-term (im)mobility, Kopczuk et al. (2010) used the rank correlation between long-term earnings in years t and $t+\tau$, where $\tau=10$, 15, 20. For each year, earnings positions are measured by the 11-year average earnings centered around the year in question. The results suggest, first, that mobility is greater the larger that T is, which is unsurprising, and yet even after 20 years, the correlation is relatively large (around 0.5 for all workers). Second, for all workers, the rank correlation decreased (mobility increased) between the early 1950s and the early 1970s and was then broadly constant. The trends differ for men from those for all workers: The mobility increase is much less pronounced and appears to rise again slightly from the early 1970s (Kopczuk et al., 2010, Figure VIII).

10.4.3 Is There More Income Mobility in the United States than in (Western) Germany?

Perhaps the most well-known "stylized fact" about income mobility is that mobility is greater in Germany than in the United States. One of the reasons for it being well known is that it is surprising: many people expect more mobility in the United States because, compared to Germany, the United States has the more flexible labor market and less comprehensive social safety net to cushion income shocks. What is often forgotten is that the original finding refers to one particular mobility concept (equalization of longer-term incomes) and to one particular time period (the 1980s, prior to German reunification in 1990).

In this section, we review the evidence about income mobility in United States compared to Western Germany. Unless stated otherwise, the data source for the United States is the PSID. We use the term "WG" to refer to the states included in the Federal Republic of Germany before reunification. The German data source, the SOEP, surveyed the former Eastern German states as well from 1990 onward, but few mobility studies to date have included these data (see later discussion). We focus on studies that examine household income mobility (which, as it happens, form the vast majority of US–WG comparative analyses). In Table 10.4, we refer to 11 studies, and summarize them in terms of the time period covered, the mobility measure(s) employed, and the main findings relevant to our question.

The pioneering study by Burkhauser and Poupore (1997) is the source of the stylized fact that we referred to earlier. It was the first major cross-comparative study of household income mobility using the new generation of comparable household panel survey data

[47] Auten et al.'s (2013) staying probabilities for years t to $t+\tau$, $\tau>1$, are also greater than the corresponding Kopczuk et al. (2010) ones because the latter's refer to presence in the top 1% in each year rather than simply in the base year and final year. For a brief discussion of persistence in the top 1% of the Canadian earnings distribution, see Saez and Veall (2005).

Table 10.4 Studies comparing household income mobility in the United States and Western Germany (WG)

Study	Time period covered	(Im)mobility measure(s)	Remarks
Burkhauser and Poupore (1997)	1983–1988	Shorrocks *R*	First finding that mobility greater in WG than in the USA
Burkhauser et al. (1998)	Year pairs t, $t+\tau$, $\tau=1, \ldots, 5$, 1983–1988	Quintile transition matrices	Slightly more income mobility in WG
Maasoumi and Trede (2001)	1984–1989	Maasoumi–Shorrocks *R*	Greater mobility in WG; statistically significant
Gottschalk and Spolaore (2002)	1983, 1993	SWF-based indices	WG–USA difference depends on index parameters
Schluter and Trede (2003)	Year pairs t, $t+1$ between 1984 and 1992	Shorrocks *R*	WG's greater mobility arises from greater mobility in low-income ranges
Van Kerm (2004)	1985, 1997	Portfolio of indices	More income movement in the USA; otherwise varies by index
Jenkins and Van Kerm (2006)	Year pairs t, $t+5$: U.S. 1981–1993, WG 1985–1999	Indices of reranking, progressivity	Reranking and pro-poorness of income growth greater in WG
Schluter and Van de gaer (2011)	Year pairs t, $t+1$ between 1984 and 1992	Index sensitive to upward structural mobility	U.S. "typically" has more mobility
Allanson (2012)	Year pairs t, $t+5$: U.S. 1981–1996, WG 1985–2004	Indices of reranking and structural mobility	Reranking and pro-poorness of income growth greater in WG
Demuynck and Van de gaer (2012)	1984–1985, 1996–1997	Indices of "inequality-adjusted" income growth	USA–WG ranking depends on weight given low-income-growth individuals
Bayaz-Ozturk et al. (2014)	5-year windows, alternating years, 1984–2006	Shorrocks *R*, ratio of permanent to total variance, log incomes	More mobility in the USA from around 1990 onward

Note: Studies are listed in order of publication year. Each study measures income as equivalized posttax posttransfer household income (using various equivalence scales), analysis samples are all individuals in households (except Burkhauser et al., 1998, all individuals aged 25–55). Western Germany: The states included in the Federal Republic of Germany before reunification.
Data sources: PSID (USA) and SOEP (WG).

becoming available in the 1990s.[48] The period covered is 1983–1988, a time of upswing in the economic cycle in both countries. Income immobility was summarized in terms of equalization of longer-term incomes using the Shorrocks R measure computed with three inequality indices (the Gini coefficient and the two Theil indices). The base year is 1983 and R is calculated as the time period is lengthened from one to a maximum of 5 years (corresponding to 1988). The headline results were summarized earlier in Figure 10.8 and refer to estimates based on the Theil index. (The other two indices yield similar profiles and orderings: see Burkhauser and Poupore, 1997, Figure 3.)

There is greater longer-term income equalization (less rigidity, lower R) in WG than in the United States in each year: the curve for the United States lies everywhere above that for WG. In numerical terms, inequality of 6-year-averaged income is 86% of average annual inequality in the United States, compared to 76% in WG (i.e., some 13% larger). The authors show that this mobility ordering is preserved if one uses different income concepts and analysis samples, including labor earnings (for all workers, workers aged 25–50, and the subsets of full-time workers in each case), and equivalized pretax pretransfer ("pregovernment") household income.[49] For example, among full-time workers aged 25–50, the 6-year R for annual labor earnings is 88% for the United States and 79% for WG. For the subset of men, the corresponding estimates are 86% and 78%; for women, 87% and 66% (Burkhauser and Poupore, 1997, Table 4).

The mobility of labor earnings over the same period is analyzed in greater detail by Burkhauser et al. (1997) using different summary methods: statistics based on quintile transition matrices, the rank correlation, and regression-based variance components modeling. Interestingly, given the subsequent focus by researchers on the US–WG differences in household income mobility, Burkhauser et al. (1997) emphasized the similarities in earnings mobility:

> *While we have found evidence of differences in the dynamic earnings movements of workers in the United States and Germany, it is perhaps the similarities of the "end results" of the two labor markets, despite substantial differences in their institutions, that highlight our multiperiod look at these two industrial giants.*
>
> ***Burkhauser et al. (1997, p. 793)***

Burkhauser et al. (1998) supplemented the two earlier studies from the Burkhauser team. As in the first study they used multiple measures of income (and associated samples), but analyzed individuals aged 25–55 years; like the second study, (im)mobility is summarized in positional terms using quintile transition matrices, not R. Again, the conclusions point

[48] Duncan et al. (1993) and Fritzell (1990) are examples of earlier cross-national studies of poverty dynamics and income income mobility using data that were not as comparable. For an earlier cross-national study of earnings mobility across 8 countries, see OECD (1996).

[49] The equivalence scale for all measures of household income in this study is derived from those in the U.S. official poverty line thresholds.

more to cross-national similarities rather than differences: "[i]ndividual mobility patterns in the two countries are remarkably similar" (Burkhauser et al., 1998, pp. 143–144). For example, the proportion of individuals in the same quintile group of posttax posttransfer household income in 1983 and 1988 is 44.7% in the United States compared to 41.4% in WG; for labor earnings mobility, the corresponding proportions are 52.6% and 53.8% (Burkhauser et al., 1998, Tables 6.2, 6.5).

It is the cross-national difference in R that received the most attention in the later studies, with most authors concerned with the robustness of the conclusion to use of different mobility indices. And all the subsequent studies that we are aware of have focused on household income, not labor earnings. Schluter and Trede's (2003) article is rather different in that they aimed to examine the Burkhauser–Poupore result in greater detail. As discussed earlier, their methodological contribution was to explain how R reflected the aggregation of distributional changes, differently weighted, at each point along the income range from poorest to richest, and to explore how the aggregation function differed by inequality index. Using a moving 2-year window over the period 1984–1992 for the calculation of R, Schluter and Trede (2003) confirmed that mobility is greater in WG than the United States. But their main substantive contribution was the finding that this difference in aggregate reflected a combination of greater mobility in low-income ranges combined with greater local weight given to these changes by the mobility index. The cross-national differences in mobility at the bottom are reminiscent of those revealed in Section 10.3 by graphical devices such the transition color plot (Figure 10.1) albeit for a different period (1985 compared with 1997).

Maasoumi and Trede's (2001) article built on earlier work by Maasoumi and Zandvakili (1986), which modified the Shorrocks R measure to use different measures of longer-term income (essentially a generalized mean rather than a simple arithmetic average). Maasoumi and Trede (2001) examined US–WG mobility differences using these Maasoumi–Zandvakili–Shorrocks indices and essentially the same household income data as Burkhauser and Poupore (1997), and also derived the sampling distribution of the indices, thereby allowing consideration of whether mobility differences were statistically significant. The substantive findings are threefold: mobility is greater for WG than the United States regardless of the indices (i.e., regardless of the measure of longer-term income, or the inequality index); that cross-national differences were statistically significant; and mobility is greatest among 16- to 25-year-olds but for all six age groups considered, mobility is statistically significantly greater in WG than the United States.

Gottschalk and Spolaore (2002) is the first (and only) paper that we are aware of that undertakes US–WG comparisons using an explicit SWF-based approach (the application considers mobility between 1984 and 1993). As indicated in Section 10.2, their approach allows for different weights to be placed on mobility as reversal and as time independence (as well as incorporating intertemporal inequality aversion of varying degrees). If the reversals and time independence aspects are ignored, so that the SWF reflects inequality-aversion considerations only, Gottschalk and Spolaore (2002) reported that the United States "gains more" from mobility than does WG. But "this reflects similar gains from reversal in the two countries but greater gains in the United States from origin independence. The introduction of aversion to intertemporal fluctuations and aversion to future risk makes the impact of mobility in the two countries more similar" (Gottschalk and Spolaore, 2002, p. 191). Put simply, conclusions about mobility differences depend on the mobility concept(s) taken and how they are weighted.[50]

Van Kerm (2004) was the first to use Fields and Ok (1999b) indices of income movement to compare the United States and WG among a portfolio of measures of household income mobility. (He also studied Belgium.) Changing the mobility concept leads to a reversal in the country ranking: the average absolute change in log incomes between 1985 and 1997 is 0.523 in the United States but only 0.392 in WG (and 0.335 in Belgium). Van Kerm remarked that "[d]ifferent concepts of mobility may indeed lead to completely different rankings of economies . . .. In all cases, mobility is higher in Western Germany than in Belgium, but the United States can stand at any of three positions depending on the index considered" (Van Kerm, 2004, p. 233). Van Kerm's decompositions highlight that the importance of distinguishing between mobility measures sensitive to positional change and those also reflecting individual income growth and changes in the marginal distributions. The "exchange" factor of distributional change is greater for WG than the United States, whereas the "growth" and "dispersion" factors are greater for the United States (Van Kerm, 2004, Table 4).

Parallel research by Formby et al. (2004) comparing mobility in individual annual labor earnings in WG and the United States between 1985 and 1990 underlines the relevance of the mobility and income concepts chosen. Using measures based on quintile transition matrices, the authors showed that there is more positional mobility in the United States than WG according to four out of five indices, and there is no dominance in the Atkinson and Bourguignon (1982) sense. However, when origin and destination earnings groups are defined as fractions of mean or median earnings (so the mobility matrices reflect real income growth as well), all five summary indices show greater mobility in the United States. The fact that US–WG positional mobility differences are less pronounced (or reversed) for individual earnings compared to household income underlines the conclusions of Burkhauser et al. (1997) cited earlier.[51]

A range of different mobility indices and time periods is used in the remainder of the studies cited in Table 10.4. Jenkins and van Kerm (2006) showed that indices of both

[50] Gottschalk and Spolaore (2002, Table A.4) provide quintile transition matrices "for comparison to other studies." Unfortunately, the scope for doing, for example, dominance checks is limited by the fact that the matrices are not bistochastic. The column sums differ greatly from 100% in several cases.

[51] The main focus of Formby et al.'s (2004) article is methodological—to derive statistical inference procedures for transition matrices and summary mobility indices based on them.

reranking and of progressive individual income growth are greater in WG than in the United States. Using related methods and data, Allanson (2012) confirmed the greater reranking in WG but also highlighted other dimensions of mobility differences. Schluter and Van de gaer (2011) and Demuynck and Van de gaer (2012) proposed classes of mobility indices that are sensitive to individual income growth, with different indices reflecting differences in the weights given to income changes of different sizes. Unsurprisingly (in the light of our earlier discussion), both papers report that mobility from this perspective is generally greater in the United States than WG but, also, the ranking can be reversed for some weighting functions.

The final article cited in Table 10.4 brings us full circle because Bayaz-Ozturk et al.'s (2014) research is in effect a reanalysis of the original Burkhauser and Poupore (1997) study, but using more up-to-date data (1984–2006).[52] The main mobility index is the Shorrocks R calculated using the Theil inequality index, but now also supplemented with estimates of the transitory variance of log income expressed as a proportion of the total variance (calculated using the Gottschalk and Moffitt [1994] "BPEA" method). If the two indices are calculated taking 1984 as the base year and extending the period over which longer-term incomes are calculated to the full 23 years (i.e., also restricting analysis to a sample with fixed structure), income mobility is greater in WG than the United States in each year. The profile of R (and for the other measure) for the United States lies above that for WG thoughout, though the gap between them gets smaller over time (Bayaz-Ozturk et al., 2014, Figure 10.1). In this sense, the results are consistent with the Burkhauser and Poupore (1997) finding (see also Figure 10.8). However, when the indices are calculated using a moving 5-year window (and hence also different samples) to examine mobility trends, an interesting result emerges, which is illustrated in Figure 10.11. We remarked earlier in an apparent increase in mobility in the United States in the late-1980s (though Bayaz-Ozturk et al., 2014 reported that the changes in their estimates are not statistically significant). Figure 10.11 shows that mobility in WG fell between the late 1980s and 1990s (the changes are statistically significant). The result is that, compared to the late 1980s when the WG–US mobility differences were statistically significant, they were no longer so in the period thereafter.

An interesting substantive question is why WG mobility fell, and to what extent it reflects changes in the (West) German labor market and economy associated with reunification or with other structural factors (observe that the downward trend apparently started before 1990). Bayaz-Ozturk et al. (2014) reported that when they applied their methods and samples to examine the mobility of labor earnings for men aged 25–59, they found similar patterns of change over time and cited Aretz (2013) as also finding a

[52] Alternating years are used to account for the change to biannual interviewing in the PSID. The authors reported that using contiguous years over the periods where it was feasible leads to similar results.

downward trend in earnings mobility when using administrative record data covering 1975–2008. Interestingly, Aretz's (2013) work shows that the downward trend in WG was broadly U-shaped between the mid-1970s and late-1980s, but did decline again sharply from around 1990. The decline in mobility the former Eastern Germany (measured only after 1990) fell even more rapidly, down to around WG levels by the mid-2000s.[53] See also Riphahn and Schnitzlein (2011), who pointed to the role of increasing job stability in Eastern Germany.

In sum, although income mobility in the United States and Germany has received much attention, there remains plenty to learn. The sensitivity of conclusions about cross-national differences suggests the need for a more comprehensive analysis using a portfolio of measures within the same study and using up-to-date data. The income concept also matters; researchers have highlighted WG–US differences in household income mobility, and the similarities in earnings mobility have received less attention. Looking at earnings mobility is also informative for tracing the sources of changes in household income mobility.

10.4.4 Intragenerational Income Mobility: Selected Other Evidence

The remainder of our discussion of evidence about intragenerational income mobility reviews cross-national comparative studies for a wider set of countries and selected country studies analyzing trends over time. The focus remains on household income mobility. We consider work done in the last two decades rather than earlier studies.

A natural place to begin is with the analysis of Aaberge et al. (2002) and Chen (2009) because both include mobility comparisons between the United States and with other countries. In the former case, the comparisons are with three Scandinavian countries (Denmark, Norway, and Sweden) in the 1980s. In the latter case, they include Canada, Germany, and GB over the 1990s. Chen (2009) also provides some information about mobility trends.

Aaberge et al.'s (2002) research is based on diverse sources of longitudinal data. For Denmark and Norway, the income data and samples come directly from registers; for Sweden, incomes refer to register data linked to respondents to the Level of Living Survey (the analysis sample is survey rather than register based), and for the United States, the source is the PSID (sample and income data come from a survey). This diversity leads to some compromises in the search for comparability. For instance, the posttax posttransfer income concept in the main analysis refers not to a household total but an aggregate across two adults (in the case of a legally married couple) or one adult (all other cases), and equivalized by the number of adults (two or one, respectively). The constraint is what

[53] This is found for both men and women and using the average jump index of positional mobility as well as the Shorrocks R.

is possible with the Swedish data: no account can be taken of cohabitation, and the number of children is unknown. As it happens, when the authors reran their analysis using more conventional definitions (but excluding Sweden), mobility levels changed for all countries, but "the mobility ordering of countries is unaffected by this sensitivity check" (Aaberge et al., 2002, p. 457).

The Aaberge et al. (2002) study provides analysis for 1986–1991 and 1990–1991, with the end chosen because a major Swedish tax reform in 1991 made later income data noncomparable (the registers covered a different combination of income sources). Mobility is measured using a Gini-based Shorrocks M index and summaries of the directional income movement in the Fields and Ok (1999b) sense. The perhaps surprising finding is that, across the four countries, and despite the substantially greater cross-sectional income inequality in the United States than the three Scandinavian countries, "the pattern of mobility turns out to be remarkably similar in the sense that the proportionate reduction in inequality from extending the accounting period for income is much the same" (Aaberge et al., 2002, p. 443). This finding arises whether the analysis is of individual labor earnings or disposable income. The "remarkable similarity" is also reported by Fritzell (1990) in an earlier study of income mobility in Sweden and the United States. Clearer cross-national differences are apparent, however, when Aaberge et al. (2002) looked at the distribution of changes in relative incomes changes between 1 year and the next over their sample period (relative income is the ratio of income to the year-specific mean; relative income change is a directional summary of individual income movement). As it happens, the distribution of relative income changes is more dispersed in the United States than in the Scandinavian countries for both individual earnings and disposable income. Once again, the conclusions about mobility that are drawn depend on the measure employed.

Chen's (2009) article is based on data from the CNEF, covering from the early 1990s to around 10 years later. Income refers to posttax posttransfer household income, equivalized using the square root scale; the analysis is of all individuals in households with positive incomes. Germany refers to the unified country here, not WG as earlier. Comparisons with the United States in the late 1990s are complicated by the move to alternate-year interviewing by the PSID.

Chen (2009) summarized short-term positional mobility in terms of 2- and 5-year IRs for decline transition matrices, calculated over moving time windows. Choice of measure matters. For example, over the 1990s, around 40% of British individuals remained in the same 10th between one year and the next, compared to nearer 50% in Canada, with Germany's rate in between. With a 5-year interval, the cross-national differences become much smaller, with the proportion remaining in the same decile group falling to between 25% and 30% for all the countries. Chen's summary refers to "a high degree of similarity in relative income mobility across nations" (Chen, 2009, p. 81), rather than to differences.

Chen (2009, Table 1) presents estimates of the Fields and Ok (1999b) index of income flux, the average absolute log-income change calculated over 5-year intervals using between 1991 and 2002. The United States and GB have broadly similar income flux over the period, Germany's is the lowest, and Canada's is in between. Only for the United States is a trend over time apparent (slightly upward). Assessment of these patterns is complicated because the estimates reflect a combination of differences in overall national income growth rates and changes in how pro-poor the income growth is. Chen (2009, Table 1) shows that economic growth accounts for an increasing share of total income flux in each country (all four countries were in an economic upswing over the period) but does not discuss pro-poorness.

Chen's final set of estimates refers to mobility as equalization of longer-term incomes, summarized using the Shorrocks measure $M = 1 - R$, with 1993 taken as the base year and time periods of up to 6 years (Canada), 10 years (GB and Germany), and 8 years for the United States (1995 and 1997 are excluded). The finding is that mobility is greatest in GB and least in Canada for all time periods, with the profiles for Germany and the United States in between and very similar to each other. Chen (2009, Figure 5) shows this for the case in which M is calculated using the mean logarithmic deviation index, but his Table A2 shows that the result is the same if calculations are done instead with the Theil or Gini index. (If half coefficient of variation squared is used, the U.S. profile is closer to Britain's.) These results echo Bayaz-Ozturk et al.'s (2014) finding of similar longer-term income equalization in the United States and WG after 1990 (see earlier). In his discussion of Burkhauser and Poupore's (1997) results, Chen commented that his results suggest that "income mobility has increased considerably in the United States between the 1980s and 1990s, while it has declined in Germany" (Chen, 2009, p. 88).

Leigh (2009) extended comparisons to include Australia, using estimates of R for periods of 2 and 3 years and using CNEF data for Britain, Germany, and the United States, plus data from the Australian household panel HILDA (HILDA data were not included in the CNEF at the time). He found that "[a]round 1990, the U.S. was more immobile than either Britain or Germany. . . . During the 1990s, Germany became somewhat less mobile, and the U.S. somewhat more mobile" Leigh (2009, p. 16) and that Australia was more mobile than all three other countries in the early 2000s.

A different set of countries is included in the cross-national analysis of Ayala and Sastre (2008), based on ECHP data covering 1993–1997: Great Britain, France, Germany, Italy, and Spain. Income is posttax posttransfer household income equivalized by the modified-OECD scale, and mobility is examined for all individuals using a balanced five-wave panel for each country. According to the Fields and Ok (1999b) index of income flux (Ayala and Sastre, 2008, Table 2), the average absolute log-income change, and looking at income changes between 1993 and 1997, Spain, Great Britain, and Italy had relatively high income flux (index values of 0.390, 0.373, and 0.360, respectively), whereas Germany and especially France were low-income-flux countries (0.309 and 0.250). Income flux is shown to be greater among individuals in single-parent households, and relatively stable among older persons (as might be expected). A second set of estimates relates to mobility as equalization of longer-term incomes assessed using the ethical indices proposed by Chakravarty et al. (1985) and calculated using multiple inequality indices and for an interval of 2 years only (individuals' base-year income is the average of their 1993 and 1994 incomes; their final year income is the average of their 1995 and 1996 incomes). Regardless of the inequality index used, Italy had the greatest mobility, but Spain slipped down the ranking and Germany rose up to second place. As the authors commented, the "results show that cross-country comparisons of income mobility can be dependent on the approach used" (Ayala and Sastre, 2008, p. 470). They also referred to potential issues related to differences in national samples (including, e.g., a relatively high attrition rate in the Spanish data), and the particular time period covered.

Gangl (2005) was more ambitious in that his mobility comparisons involved eleven EU countries (data from the ECHP) and the United States (PSID). The periods covered are 1994–1999 (ECHP) and 1992–1997 (PSID). Income is equivalized posttax posttransfer household income samples restricted to individuals aged 25–54 years. Gangl calculated two principal measures, namely, Shorrocks *R* for a 6-year period and the transitory variance of log income expressed as a proportion of total inequality (derived using a regression decomposition). Discussing *R*, Gangl emphasized similarities across countries rather than differences: For example, using a Theil-based index, "about 75–80% of observed income inequality has been permanent over the 6-year observation period in most countries" (Gangl, 2005, pp. 149–151). Nonetheless, Germany, Ireland, and the United States are relatively immobile countries and the Netherlands and Denmark the most mobile ones. Interestingly, "low-inequality countries . . . also tend to be the countries exhibiting the lowest degree of persistence in income inequality over time" (Gangl, 2005, p. 151). Germany is an exception to this description: It is a relatively low inequality country but also with relatively high immobility. This description of Germany also fits with the findings of Aaberge et al. (2002) discussed earlier (for an earlier period). In sum, and on balance, it is unclear whether there is a positive relationship between cross-sectional inequality levels and rigidity of longer-term incomes.

Gangl's (2005) results for household income are consistent with those of Gregg and Vittori (2009), who examined the mobility in labor earnings of individuals aged 20–64 in Denmark, GB, Germany, Italy, and Spain, also using ECHP data. Using *R* calculated for different inequality indices, they found that longer-term earnings inequality reduction is greatest in Denmark followed by Italy, and Germany is the least mobile, with GB and Spain in between. Applying the methods of Schluter and Trede (2003), Gregg and Vittori (2009) also found that most of the cross-national mobility differences are accounted for by differences in mobility patterns in the lowest earnings ranges.

With his variance components measure, Gangl (2005) found that

> *most (i.e., 65%–70%) of the observed total income inequality for any single country is permanent income inequality with countries like Denmark, the Netherlands, Spain, or Italy at the low end and Ireland, Portugal, the United States, and Germany at the upper end of the scale. Still, cross-national variations in relative income persistence are small, and the country ranking in terms of permanent income inequality in consequence almost exactly mirrors the country ranking for total income inequality*
>
> **Gangl (2005, p. 152)**

However, if one focuses on the variance components in absolute levels rather than as expressed as a share of the total variance, the picture changes somewhat. For example, the countries with the lowest transitory variances are Denmark, Germany, and Ireland, and the largest are for Italy, the United States, and Spain.[54]

The most comprehensive analysis of income mobility to date using the new EU-SILC longitudinal data is by Van Kerm and Pi Alperin (2013), who also pointed to a number of important issues concerning the cross-national comparability of the constituent data sources and the short period covering by the data. On these issues, see also Jenkins and Van Kerm (2014).

We now turn to country studies of income mobility with a focus on trends, of which there are few. Jenkins's (2011a) book contains a comprehensive study for GB, using BHPS data covering from 1991 through to 2006 and examining trends in various concepts of mobility.[55] The headline finding is that, for all but one concept of mobility, there is virtually unchanged mobility throughout the period. This is found for a portfolio of measures, including 1-year positional mobility, Shorrocks R measures calculated over moving 6-year windows, and the transitory variance of log household income (Gottschalk and Moffitt, 1994, "BPEA" method, using moving 7-year windows).[56]

Jenkins (2011a) reported the same lack of trend if one looks at the earnings of prime-age men and women: see also Jenkins (2011b) and Cappellari and Jenkins (2014). (These studies also cautiously suggest that transitory variances of household income and men's earnings in Britain are larger than their counterparts for the United States.) The lack of

[54] The estimates are not directly comparable with those reported for the United States earlier because Gangl's (2005) decomposition uses a model specification that is nonstandard. For example, he did not allow for persistence over time in transitory shocks. Also, Gangl included the variance of heterogeneous income trends in the transitory component rather than permanent component.

[55] The book also reviews earlier British studies of income mobility, most of which are by Jenkins and collaborators and based on shorter spans of BHPS data. For example, Jarvis and Jenkins (1998) used four waves; Jenkins (2000) used six waves.

[56] Using Jenkins's (2011a) data, we have compared the decile transition matrices for 1991–1998 and 1999–2006 and found that there is no stochastic dominance. Nor is there if we compare the British matrix for 1991–1998 with the decile transition matrix for the United States for 1989–1998 shown in Table 10.1.

change in earnings mobility during the 1990s found in BHPS data was also found by Dickens and McKnight (2008) using administrative record data on earnings covering the period between financial years 1978/1979 and 2005/2006.[57] They summarized mobility using the Shorrocks equalization measure $M=1-R$, calculated using multiple inequality indices, and over moving windows of 2, 4, 6, 8, and 10 years, and each series tells the same story. Interestingly, Dickens and McKnight's (2008) research also found that mobility was on a downward trend between 1978/1979 and the beginning of the 1990s (though this trend is less pronounced for women than for men).

Jenkins (2011a) observed that the lack of change in British income mobility between the early-1990s and mid-2000s is surprising given significant changes over the period considered in tax–benefit policies and the upswing in the macro-economy from trough to peak. Jenkins (2011a, chapter 6) adduced some evidence to suggest that the lack of trend in aggregate may reflect a balance between changes in mobility associated with different income sources comprising total household income, but he conceded the exploratory nature of the analysis.

The exceptional measure for which some (relatively small) changes are observed is in the pattern of individual growth. Jenkins and van Kerm (2011) showed that income growth between 1998 and 2002 was more pro-poor than in earlier periods (1992–1996 and 1995–1999), but not so compared with 2001–2005. (An extract from their results was shown earlier in Figure 10.7.) The authors suggested that the pro-poor nature of individual income growth in the 1998–2002 period arose because the economy was buoyant, with unemployment rates continuing to fall relatively rapidly from their early-1990s peak, and the incoming Labour government had an explicit antipoverty agenda, unlike the preceding Conservative governments. It is speculated that the subsequent fall in the progressivity of income growth had to do with the slowdown in the economy from around 2000.

Trends in transitory (and permanent) earnings variances of earnings in Western Germany were studied by Bartels and Bönke (2013). Bartels and Bönke worked with samples of man aged 20–59 years over the period 1984–2009, calculating variance components using the Gottschalk and Moffitt (1994) "BPEA" method (using moving 5-year windows). The striking finding (2013, Figure 2) is that, although the transitory variance of log earnings rose over the period as a whole, the transitory variance for equivalized posttax posttransfer household income (for the same sample) does not change at all over the period, pointing to important roles played by the German welfare state and by families in offsetting shocks to men's earnings. When the same methods were applied to Britain (BHPS data for 1991–2006), Bartels and Bönke (2013) found, like Jenkins (2011a), that

[57] They make pioneering research use of the Lifetime Labour Market Database (LLMDB), a 1% sample of individuals identified by National Insurance numbers and originally designed to estimate workers' National Insurance contributions and State retirement pension entitlements.

the transitory variance for equivalized posttax posttransfer household income did not change over time; unlike him, they also reported (2013, Figure 6) a rise in the transitory difference of men's earnings (and higher levels). These differences are traced to differences in samples: Jenkins (2011a) considered men aged 25–59 (as in most similar U.S. studies), whereas they argued that transitory earnings shocks are more important for this group. Overall, the authors concluded from their analysis that "redistribution and risk insurance provided by the welfare state is more pronounced in Germany than in the United Kingdom" (Bartels and Bönke, 2013, p. 250). Whether this also applies to other groups beyond prime-aged men requires examination.

Mobility in top incomes in Germany over the period 2001–2006 was studied by Jenderny (2013) using tax administrative data, a 5% balanced sample of all tax filers in those years. Income is the tax unit's gross pretax income (i.e., including tax-exempted income, but not realized capital gains). One-year probabilities of remaining in the top 1% are about 78% and thus larger than the estimates of around 70% reported by Auten et al. (2013) for nonrecessionary periods in the United States (see earlier discussion). Five-year survival rates are also larger in Germany than in the United States.[58] Jenderny (2013, 32) concluded that the increase in top income concentration in Germany since the 1990s described by Bach et al. (2009) is unlikely to be offset by high or rising top income mobility.

10.4.5 Summary and Conclusions

Empirical studies of income mobility show that, in all countries, there is a substantial degree of longitudinal flux in incomes, whether looking at incomes 1 year apart, or 5 or 10 years apart, resulting in changes in relative position and a reduction in the inequality of longer-term incomes. It is also clear, however, that most income changes are relatively small so that, even after many years, relative positions are quite highly correlated and substantial inequalities in longer-term incomes remain.

To the big questions of whether income mobility in country A has increased or decreased over time, or is greater or less than in country B (or C or D or . . .), we have found few clear-cut conclusions—apart from a general finding that the answers to the questions depend on the mobility concept that is used, and other issues such as the time period considered and the measure of income are relevant.

This is illustrated by the comparisons of the United States with WG. Early research suggested that income mobility in the 1980s was (surprisingly) greater in WG than in the United States (Burkhauser and Poupore, 1997) when mobility is measured in terms of equalization of longer-term income. But more recent research (Bayaz-Ozturk et al.,

[58] The stayer rates for the top 0.1% are also slightly higher than the Canadian estimates reported by Saez and Veall (2005, Figure 2).

2014) for the 1990s using the same measure suggests that mobility in the two countries is now similar. And it is often forgotten that the Burkhauser team had long argued that earnings mobility in WG and United States was remarkably similar. Moreover, when one switches the mobility concept to one of income movement (or individual income growth), mobility in the United States shows up as greater than in most other countries—the ranking consistent with many people's expectations given the nature of the U.S. economy, labor market, and welfare state.

It remains an open question, as well, whether there is a systematic cross-national relationship between levels of income mobility and cross-sectional income inequality. The evidence is mixed, and the issue deserves to be revisited. (Note the widespread interest too in whether there is a corresponding relationship for intergenerational income mobility—see the discussion of the "Great Gatsby" curve in Section 10.5). Because the evidence we have reviewed suggests similarities across countries in the extent of mobility (positional and longer-term income equalization) rather than marked differences, we are inclined to conclude that there is no obvious relationship between mobility and inequality because cross-national differences in inequality are pronounced.

Looking at trends over time in income mobility within countries, the picture is one of diversity and depends on the mobility concept and the length of time period over which trends are assessed. Mobility changes are observed in the United States over the 30 years since the early 1970s and in Germany between the late 1980s and the 1990s, though whether these count as large or small changes partly depends on the eye of the beholder. For Britain, there is a clearer case that income mobility in Britain changed hardly at all in the 1990s and 2000s (again with the exception of mobility as individual income growth). Relatively large changes in mobility are more apparent to most eyes once trends are assessed over a relatively long period. The U.S. study of earnings mobility by Kopczuk et al. (2010), with data going back to 1937, is the best example we have of this.

In sum, our review of evidence about income mobility suggests that there is much to learn. The advent of cross-nationally comparative household panel surveys over the last three decades facilitated a relative boom in intragenerational mobility analysis. There are signs that the next generation of studies will make greater use of administrative register data or surveys linked to administrative data, at least for analysis of trends over time. As we have discussed, data from sources such as tax administrative records provide the advantages of huge samples with good coverage of top incomes and can provide long historical series as well. On the other hand, these benefits come at the potential costs of having income definitions that are not as useful for mobility analysis as those now in comparative survey collections such as the CNEF (and may change over time as tax laws change), and data access and undertaking the analysis are also nontrivial issues. For cross-national comparisons, administrative record data also have potential, but the problems of comparability are an order of magnitude greater, and data may simply be unavailable for countries of key interest.

10.5. INTERGENERATIONAL MOBILITY: EVIDENCE

Section 10.3 presented a set of measures by means of which we can describe not only intra- but also intergenerational associations in a society. This section reviews evidence on such associations.

There are several earlier reviews of intergenerational income mobility. Solon (1999) reviewed intergenerational labor market with a focus on long-run earnings, whereas Solon (2002) focused on a subset of that literature, namely cross-national differences in mobility. Björklund and Jäntti (2009) built on and extended the empirical evidence assembled by Solon (1999). Black and Devereux (2011), who examined intergenerational links in income and education, emphasized evidence on causal links in intergenerational mobility. Blanden (2013) contrasted the crossnational evidence on intergenerational income, earnings, and education mobility with mobility in social class. Corak (2006, 2013a), in turn, emphasized policy implications. Corak (2013a) also drew on recent research about both socioeconomic gradients in child development and the emergence of economic persistence in labor markets.

Several recent reviews present international evidence on intergenerational income persistence in a scatter plot, plotting the estimated persistence in different countries on the vertical axis and estimated income inequality, often in the parental generation, on the horizontal, adding a linear bivariate regression line (Björklund and Jäntti, 2009; Blanden, 2013; Corak, 2013a). Labeled the "Great Gatsby" curve by the then-chairman of the U.S. Council of Economic Advisors (Krueger, 2012), such plots are interpreted to suggest countries with higher persistence are also countries with greater inequality. Figure 10.13 reproduces the most recent such graph, from Corak (2013a, Figure 1). Although the precise estimates used by different authors vary, the results are broadly similar. The Nordic countries have low persistence and low inequality; the United States, the UK along with France and Italy, have high persistence and reasonably high inequality.

There are theoretical models that can account for the positive association between inequality and persistence. For instance, in Solon's (2004) version of the Becker and Tomes (1979, 1986) model, the factors that drive intergenerational persistence, such as the heritability of human capital endowments, the returns to education, and the progressivity of public education expenditure, affect cross-sectional inequality with the same signs. In Hassler et al.'s (2007) model, which examines links between inequality and mobility under different kinds of labor market institutions, some institutional arrangements have mobility and inequality being inversely related (and hence persistence and inequality positively correlated). Checchi et al.'s (1999) model of beliefs about own ability, educational choice and mobility can also generate positive as well as negative associations between inequality and mobility depending on the model parameters. As we shall see, however, it is far from clear that intergenerational persistence and inequality are, in fact, as clearly positively correlated as Figure 10.13 suggests.

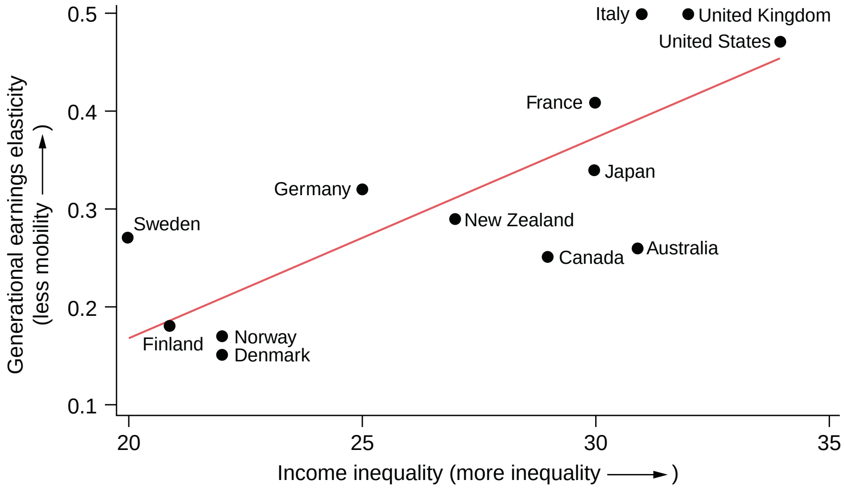


Figure 10.13 The Great Gatsby curve: the relationship between intergenerational earnings persistence and cross-sectional income inequality. *Note: Income inequality is measured by the Gini coefficient of disposable household income in 1985 taken from the OECD. Persistence is measured as the Beta of parental and son earnings. Sons are born in early 1960s, and outcomes for them are measured in late 1990s. See Corak (2013a,b) for further detail. Source: Corak (2013a, Figure 1).*

This part of the chapter proceeds as follows. In Section 10.5.1, we discuss data requirements and special problems that come up in estimating intergenerational and family associations. In Section 10.5.2, we review studies of intergenerational persistence and mobility in the United States. The focus on this country is motivated, as in the intragenerational mobility case, by the sheer amount of evidence about mobility in the United States relative to that in other countries. First, we examine evidence on the level of the intergenerational elasticity (IGE) of earnings or income—first for father–son pairs, and then widen the scope to look at broader pairings of parents and offspring—and then examine evidence about trends in the IGE over time. (The IGE is the Beta measure discussed in Section 10.3; we use both terms interchangeably in this section.) We then examine evidence that is based on measures that go beyond the simple log-linear Galtonian regression Beta (IGE), product–moment correlation coefficient *r*: for example, quantile regressions, transition matrices, nonparametric conditional mean functions. In Section 10.5.3, we examine evidence on intergenerational mobility from other countries, following the same structure as for the United States. In Section 10.5.4, we examine evidence on another way to measure the importance of family background, the sibling correlation; and in Section 10.5.5, we discuss other approaches to intergenerational mobility, old and new. Section 10.5.6 concludes.

10.5.1 Data and Issues of Empirical Implementation

As discussed in Section 10.4.1, any study of income mobility faces three "W" issues: mobility of What, among Whom, and When? For intergenerational mobility, each question must be answered twice, once in each of the parental and offspring generations. As with intragenerational mobility, researchers' choices are constrained with the available data.

At one level, just as with intragenerational income mobility, mobility of "What" refers to the income concept that is used. The overwhelming majority of studies we review use the labor market earnings of the parent and the offspring with several variations, discussed in Section 10.4.1. Other choices might add nonlabor income sources from the market such as capital income to measure factor or market income. If the goal is to examine the intergenerational association of living standards, it would make sense to study disposable income (i.e., to add public transfers and deduct income taxes paid). It would seem reasonable to have identical answers to the "What" question in both generations. It is frequently the case that available data do not support such a choice; it is not unusual for "income" to be family income in the parental generation and to be earnings in the offspring generation.[59]

The aim of early research on this topic was to measure the intergenerational association of "permanent" income, which was believed to be captured quite well by labor market earnings. It has long been recognized (see Atkinson, 1981b) that short-run income measures are different from longer-run measures because of transitory fluctuations, and that the associations sought were those of the more stable or permanent measures of living standards.

As with intragenerational mobility, "Whom" refers to the definition of the income-receiving unit in both the parent and child generation. Most studies that are modeled on Solon (1992) examine mobility of father–son pairs, ignoring the incomes of other household members. Many departures from this are due to data-related reasons. For instance, studies such as that of Zimmerman (1992), which relies on data from the U.S. National Longitudinal Survey of Youth (NLSY), uses family income in the parental generation, as that is the only income concept available in that data source.[60]

The Whom question becomes more complex when the intergenerational associations of women's incomes are studied and compared with those of men. Over the last four to five decades, women's labor market attachment has increased substantially in most developed nations, with female labor force participation rates increasingly resembling those of men. However, around the age commonly believed to be appropriate for measuring men's long-run income (around age 40), women often have breaks from employment

[59] This is the case in U.S. studies that rely on the NLSY and UK studies that rely on the BCS and NCDS. Note that the Galtonian regression of child height on parent height also used midparent height on the right-hand side; see Galton (1886) and Goldberger (1989).

[60] The Galtonian regression Beta that is the most often used measure of (im)mobility originally related offspring height to "midparent" height. See Goldberger (1989) and Galton (1886).

due to childbirth and child care. Studies that examine women's intergenerational mobility are more likely to examine family or household income as a better gauge of their living standard than individual incomes. Comparing mobility across men and women would then naturally also need to examine family or household income for men (Chadwick and Solon, 2002; Raaum et al., 2007).

There is an added dimension to the Whom question, namely the nature of the parent–offspring relationship. In the early intergenerational studies of Atkinson (1981b), Solon (1992), and Zimmerman (1992), the parent–child association was more or less driven by the survey design—"children" were the children of the sample parents who were followed up in adulthood. However, children can have multiple parents—stepparents, adoptive and foster parents in addition to birth parents. Common choices are to restrict the population to those parent–offspring pairs where the offspring was observed as living with the parent at some age, say 10 or 16, or to birth parents. One aspect of this Whom dimension is the role of separated families. Should one focus on associations of offspring income with the head in lone-parent families or on father–child associations? Some studies, especially based on register data, have examined the sensitivity of the population of parent–child relationships and found differences across definitions and family types to be relatively small.[61]

As with the two other "W" questions, the When question for intergenerational mobility analysis is mostly a superset of that for intragenerational mobility. Most of the same questions addressed in Section 10.4.1 need to be resolved for both the parent and offspring generations. The underlying data record income for a specific period: often annual income data but in some cases "current" income data are available. But, in contrast to the case of intragenerational mobility, shorter-run fluctuations are noise that make more difficult the uncovering of the more interesting underlying longer-run incomes. This leads directly to the issue of over what periods, and over what ages, incomes should be studied (and aggregated) to give reasonable measurements of longer-run economic status. And if, due to data limitations, ideal measurements cannot be made, how are mobility measurements affected? The two main issues that have been addressed are transitory variation in observed income measures, and "life cycle" bias (Jenkins, 1987; Grawe, 2006). We discuss these in turn.

Since at least Atkinson (1981b), it has been recognized that transitory errors in parental income lead to an errors-in-variables (downward) inconsistency in the estimated intergenerational elasticity. Since the seminal paper to empirically address this issue (Solon, 1992), many studies have exploited this finding.[62] Solon's estimate of intergenerational persistence for the United States, based on averages across 5 years of parental

[61] See Björklund and Chadwick (2003), and, using mobility in education, Holmlund et al. (2011) and Björklund et al. (2007b).

[62] It is much less common to correct for measurement error in other ways to asses mobility, such as transition matrices. For the econometrics involved and evidence based on simulations, see O'Neill et al. (2007).

income, resulted in point estimates of Beta that are between 10% and 70% larger than the estimates derived using a single year of parental income.

Recent work on so-called generalized-errors-in-variables (GEIV) model calls into question the assumption that transitory income variations have the same properties as classical measurement errors (Böhlmark and Lindquist, 2006; Haider and Solon, 2006). The GEIV model for the annual income process of an individual in family i in generation j (=Offspring, Parent) at age t relates permanent income y and transitory errors v to annual or current income by (Haider and Solon, 2006)

$$y_{ijt} = \lambda_{jt} y_{ij} + v_{ijt}, \quad j = O, P. \tag{10.16}$$

The key advance here is the introduction of the age-dependent parameter λ, which "loads" underlying permanent income onto annual income and is hypothesized to be lower than one early in the life cycle, equal to one at some point, and higher than one thereafter. Note that we allow for the λ parameters to differ across generations.[63]

The measurement error model in Equation (10.16) is the same as the classical measurement error model if (i) $\lambda_{jt} \equiv 1$, and (ii) the random fluctuations v are orthogonal to true long-run income ($y \perp v$), and the vs are identically and independently distributed within a generation. An estimate of the IGE β using annual incomes for both parents and children has the probability limit

$$\begin{aligned} \text{plim}\hat{\beta} &= \frac{\text{Cov}[y_{iOt}, y_{iPt}]}{\text{Var}[y_{iPt}]} \\ &= \frac{\text{Cov}[y_{iO}, y_{iP}] + \text{Cov}[v_{iOt}, y_{iP}] + \text{Cov}[y_{iO}, v_{iPt}] + \text{Cov}[v_{iOt}, v_{iPt}]}{\text{Var}[y_{iP}] + \text{Var}[v_{iPt}] + 2\text{Cov}[y_{iP}, v_{iPt}]}. \end{aligned} \tag{10.17}$$

For the classical measurement error model (and assuming, in addition, that the random fluctuations v are uncorrelated across generations), the last three terms in the numerator in Equation (10.17) are all zero. In that case, also the third term in the denominator is zero, and only the presence of the random fluctuation in parental income in the denominator leads to downward inconsistency, a point made in this context first by Atkinson (1981b). Denote the variance of permanent earnings in generation $j = 0, P$ by $\text{Var}[y_{ji}] = \sigma^2_{y_j}$ and similarly the variance of transitory earnings by $\text{Var}[v_{jit}] = \sigma^2_{v_j}$. The most common empirical solution to deal with the inconsistency is to diminish it by taking multiyear averages of parental income, in which case the measurement error variance in the denominator is $\sigma^2_{y_P}/T < \sigma^2_{v_P}$, an approach first used by Solon (1989, 1992) and Zimmerman (1992) and now standard in the literature.[64]

[63] The current exposition treats the transitory errors as being white noise. In case they are autocorrelated, the attenuation factor involves also the parameters of that process. If v follows an AR(1) process, such autocorrelation worsens the errors-in-variables inconsistency. See, for example, Mazumder (2005b, Table 1).

[64] A consistent estimator can also be constructed if there is a valid instrument for permanent income, an approach that has also been used (see, e.g., Dearden et al., 1997).

However, the age- or time-dependent factor loading λ_{jt} leads to two additional sources of bias in the IGE, namely the age/time point at which child incomes is measured—leading to a biased estimate of Cov[y_{iO},y_{iP}]—and when parental income is measured—leading to biased estimates of both Cov[y_{iO},y_{iP}] and Var[y_{iP}].

If only parental income is measured with error, we would have

$$\text{plim}\hat{\beta} = \frac{\text{Cov}\left[y_{Oit}, y_{Pi}\right]}{\sigma^2_{y_P}} = \theta_{Ps}\beta, \tag{10.18}$$

where s is the age at which parental income is measured, and

$$\theta_{Ps} = \frac{\text{Cov}\left[y_{Pis}, y_{Pi}\right]}{\text{Var}\left[y_{Pis}\right]} = \frac{\lambda_{Ps}\sigma^2_{y_P}}{\lambda^2_{Ps}\sigma^2_{y_P} + \sigma^2_{v_P}} \tag{10.19}$$

is the linear projection coefficient of y_{Pi} on y_{Pis} (Haider and Solon, 2006). If parental income is measured at an age at which $\lambda_{Ps} \approx 1$, θ is the standard errors-in-variables attenuation factor. If only *offspring* income were characterized by the GEIV process, the probability limit of the IGE estimated using annual income would be

$$\text{plim}\hat{\beta} = \frac{\text{Cov}\left[y_{Oit}, y_{Pi}\right]}{\sigma^2_{y_P}} = \lambda_{Ot}\beta. \tag{10.20}$$

Whenever $\lambda_{Ot} \neq 1$, the IGE is inconsistently estimated (Haider and Solon, 2006). Thus, when in the life cycle offspring incomes are measured matters for estimates of intergenerational persistence. This is the issue of "life cycle bias."

If both offspring and parental incomes are characterized by the GEIV model, which is plausible, the estimated IGE is (Gouskova et al., 2010; Haider and Solon, 2003)

$$\text{plim}\hat{\beta} = \lambda_{Ot}\theta_{Ps}\beta. \tag{10.21}$$

Note that λ can be either below or above unity, and θ is constrained to lie in the unit interval, so β can be underestimated (when $\lambda < 1$), be correctly estimated (when $\lambda_{Ot}\theta_{Ps} \approx 1$), and be overestimated (when $\lambda_{Ot}\theta_{Ps} > 1$). Finally, the Pearson correlation coefficient r has in this case the probability limit

$$\text{plim}\hat{r} = \theta_{Ot}\frac{\sqrt{\lambda^2_{Ot}\sigma^2_{y_O} + \sigma^2_{v_O}}}{\sigma_{y_O}}\theta_{Ps}\frac{\sqrt{\lambda^2_{Ps}\sigma^2_{y_P} + \sigma^2_{v_P}}}{\sigma_{y_P}}r. \tag{10.22}$$

The probability limit of the correlation coefficient depends on the θ in both generations, on the ratio of the observed standard deviation to that of long-run income in both generations, as well, of course, on the true r.

Empirical evidence from both the United States and Sweden on the age profile of λ_t based on the GEIV model, suggests that earnings early in life (even abstracting from a population age-earnings profile) are a downward-inconsistent measure of lifetime earnings and later in life an upward-inconsistent measure (Böhlmark and Lindquist, 2006; Haider and Solon, 2006). Around age 40, at least for men in both the United States and Sweden, $\lambda_t \approx 1$ in which case deviations from a multiyear average are approximately classical, thus lending themselves to the analysis of intergenerational association of long-run income under the assumption that the λs in both generations are approximately equal, i.e., $\lambda_{Pt} \approx \lambda_{Ot}$ (or at least that they were equal at about the same age).

Grawe (2006), building on insights in Jenkins (1987), examined the extent of both attenuation and life cycle bias in Betas estimated in several different countries and data sets. He found, using data for Canada, Germany, and the United States, that life cycle biases in fathers' age are an important source of bias and proposed several rules of thumb to diminish it: either to use points in time at which measurement errors are roughly classical, as suggested earlier, or at least to use observations on income for parents and children at similar points in their life cycle.

There are several caveats, however. First, as implied earlier, the λs may well change from one generation to the next. Second, the λs that apply to, say, earnings, may differ from those that apply to, say, disposable household income. Third, the λs that apply to men may be quite different than those that apply to women depending, for instance, on patterns of labor force withdrawal and reentry due to child bearing. Finally, the λs can be quite different in different countries. Without access to estimates of these, cross-country differences in the IGEs can be driven not by differences in the underlying βs but by different values of λ_{Ot} and θ_{Ps}, even if the ages at which incomes are measured are kept constant.[65]

It may be interesting to know how large the bias in Beta or r is in a given population. However, we are often interested in comparing these parameters in two different populations, for example, across time in a country, or between two countries. Denoting the two populations by A and B and focusing on Beta, and assuming for simplicity we are measuring both parents and offspring at the same ages, we have

$$\hat{\beta}^A - \hat{\beta}^B \simeq \lambda_{Ot}^A \theta_{Ps}^A \beta^A - \lambda_{Ot}^B \theta_{Ps}^B \beta^B. \tag{10.23}$$

Unless we have estimates of λ and θ in both countries, we must assume that $\lambda_{Ot}^A \theta_{Ps}^A \approx \lambda_{Ot}^B \theta_{Ps}^B$ is able to deduce from the estimated difference that the underlying βs are different, also. A similar argument applies, of course, to r. We can in that case infer

[65] Moreover, Nybom and Stuhler (2011) used nearly complete actual lifetime incomes for both fathers and sons. By comparing regression coefficients based on multiyear averages of sons' income with that based on their full lifetime incomes, they found that the biases in the intergenerational elasticity estimates are still quite considerable. This may mean that more complex models that link short-run to "permanent" income need to be explored.

the sign of their difference without bias (but cannot know its size unless we know λ and θ), or we can estimate their ratio.

The almost exclusive focus on permanent income (which, in some sense, involves both the What and When questions) can be questioned in light of the more complicated measurement models that link short-run to long-run income. The focus on permanent income is based on the notion that differences between short- and long-run income are transitory and largely classical, i.e., positive and negative shocks are roughly as likely (low or not autocorrelation), and the magnitude of the shocks does not vary by either permanent income or other characteristics (shocks are homoscedastic and orthogonal to permanent income). If capital markets are well functioning and individuals have a fair idea of what their permanent income is (so they know if they have been hit by a negative or positive shock), they smooth their consumption by relying on saving and borrowing. These demanding conditions would justify the focus on permanent income (see the discussion in Section 10.2). In this view, it is permanent income, not short-run fluctuations, that best captures the distribution of well-being.

It follows that, if the assumptions are violated, even short-run fluctuations are interesting from a well-being perspective. Jäntti and Lindahl (2012) demonstrated that income volatility in Sweden is strongly but nonmonotonically associated with the level of long-run income. Moreover, analysis of intergenerational associations in income suggests that not only long-run incomes but also income volatility is associated across generations (Jäntti and Lindahl, 2012; Shore, 2011). Thus, a focus on long-run incomes alone probably understates the extent to which economic well-being is associated across generations.

Before we discuss commonly used data sources for intergenerational analysis, we point to an additional complication in interpreting the evidence. The most commonly used measure of intergenerational mobility is the persistence as measured by Beta (the IGE). Arguably, we would like to abstract from the marginal distribution of offspring and use the correlation, r, related to Beta by the ratio of parental to offspring standard deviation (see Equation 10.4). In steady state, the two are equal but, when inequality increases (decreases) across generations, r is lower (higher) than Beta. Extra care should then be taken in comparing Betas across countries, as different Betas may be consistent with the same, or at least more similar r, depending on how marginal distributions have changed across generations in the two countries. Björklund and Jäntti (2009) cited evidence that suggests the ratio σ_P/σ_O in the United States is less than one—inequality increased—and in Sweden greater than one—inequality decreased—suggesting the difference in r is likely to be less than that in the Betas. (Note, however, that it is only changes in inequality in the marginal distributions that are controlled for, and in a particular way.)

Suitable data for intergenerational analysis need to meet two basic criteria. The data need to be able to identify and link parent–child pairs.[66] They need to include measurements of the incomes in both generations at comparable points in the life cycle and preferably multiple such measurements to allow for the effect of measurement errors.

Three main types of data are used. Many studies rely on longitudinal household surveys that have been running long enough to allow for the offspring to be observed living with their parent(s) as a child or youth and then followed up as adults, having often formed households of their own. These data sources, discussed in greater detail in Section 10.4.1, include the U.S. PSID and the German SOEP. The UK BHPS has only recently been used for intergenerational analysis. Cohort studies are another type of data that are commonly used for intergenerational analysis. Such data sets, including the U.S. National Longitudinal Study of Youth (NLSY) and the UK National Child Development Study (NCDS; cohort of 1958) and British Cohort Study (BCS, cohort of 1970), have been specifically designed to collect data on children and follow them across time as they grow older. The income and other information used are mostly gathered through interviews with the parents and children in such studies, and the parent–child link is ascertained from both information about both birth and living arrangements.

A variation of the survey-based approach was taken in the UK, where Atkinson (1981b, 1983), using data originally collected by Rowntree and Lavers (1951) for the study of cross-sectional poverty in York, built an intergenerational data set by interviewing the adult children of the original survey households, creating a longitudinal data set from what was originally a cross-sectional data set.

Register-based data sets are another important source of data. Such data, which underlie intergenerational mobility estimates in Canada and the Nordic countries, and increasingly also in the United States, rely on administrative records, often drawn from data originally collected for purposes of taxation or Social Security, to measure income, and identify parent–child links based either on administrative records that link parents to children or on census data.[67] The key to the use of such data is the use of personal identifiers and the presence of a reliable parent–child link.

A third approach to data is to use synthetic parent–child links. One way to do this is to use two-sample methods (i.e., to estimate Beta using empirical moments based on different data sets). This requires a sample of "parents" to provide information on the

[66] Intergenerational persistence can be estimated using two-sample methods (Björklund and Jäntti, 1997), which we discuss later.

[67] For instance, Canadian father–son pairs in studies such as Corak and Heisz (1999) rely on tax records for not only the earnings information, but also the father–son link. In the Nordic countries, parent–child links are from either Census data or from birth records (Bratsberg et al., 2007).

unconditional distribution of income in the parental generation and of the distribution conditional on a few key predictors of income, as well as a sample of "offspring" to provide information on both their income distribution and on the predictors for their parents. Two-sample methods, first used in the intergenerational context by Björklund and Jäntti (1997) for a comparison of the United States and Sweden, have later been used in several countries, including Britain, Italy, France, Brazil, and Australia.[68]

Each of these three types of data is subject to measurement challenges (see Section 10.4). Measurement errors in income and other data used in the analysis are an issue, and not only in survey- but also in register-based sources, although the nature of the errors are likely different (e.g., recall error in survey data and underreporting due to tax evasion in register data).[69] Attrition, especially selective attrition, is a concern in longitudinal surveys, a problem that may be compounded in intergenerational follow-up. The reliability of the identification and linking of parents and children is a concern when that is done using administrative data.

Before we delve into the evidence, we should note that the overwhelming majority of studies, in the United States as well as in other countries estimate elasticities (i.e., estimates of the Beta measure discussed in Section 10.3). When correlations are available (either directly, reported by the authors, or derived using Equation 10.4 based on Betas and the standard deviations), they are product–moment (Pearson) rather than rank (Spearman) correlations. Fully controlling for the marginal distributions would require the latter. Moreover, we are unaware of any study that explicitly recognizes the implications of the GEIV model for estimated elasticities that attempts to control for those effects.

One reason most analysts may estimate Betas rather than Pearson or Spearman correlation coefficients, or transition matrices, is convenience: controlling for systematic life cycle effects in both generations is simple in a multiple regression framework. Moreover, the impact of transitory errors, when classical, is well understood and simple to mitigate. Estimation of the Pearson correlation is subject to the same errors-in-variables inconsistency as the Beta, but transitory errors in both offspring and parent income cause r to be underestimated, so reducing the inconsistency requires time-averaging income in both generations. Transitory errors lead to inconsistent estimates of rank correlations also. O'Neill et al. (2007), who presented simulation evidence based on bivariate normal distributed parent–offspring income that are subject to a range of different kinds of measurement error, suggest robustly that intergenerational persistence is underestimated and mobility overestimated in the presence of measurement error. Finally, in many cases,

[68] Two-sample methods were independently developed by Angrist and Krueger (1992) and Arellano and Meghir (1992). Methods to estimate the variance of such estimators were derived by Inoue and Solon (2011).

[69] For a discussion, see, for example, Ehling and Rendtel (2004).

Beta and *r* are estimated using instrumental variables, often using sample moments from different samples. These techniques are well understood in moment-based estimation, but less so for rank correlations and nonparametric techniques.

10.5.2 Intergenerational Persistence in the United States

Although there are many studies of intergenerational mobility in the United States as well as in other countries, the literature is characterized by a surprising number of omissions. For instance, we have been unable to locate transition matrices for different cohorts of parent–child pairs, so we are unable to examine the change across time in mobility using the dominance approach.[70] Most U.S. researchers report only Betas, not *r* or rank correlations, so standardization for changes in the marginal distribution of earnings or income is incomplete at best. And yet this is a period in which there have been pronounced increases in inequality in the United States (and many other countries).

By the late 1980s, two longitudinal data sets in the United States, the Panel Study of Income Dynamics (PSID) and the National Longitudinal Study of Youth (NLSY) had been running for sufficiently long to allow the study of the incomes of parents and children at economically active ages. Around that time, three papers were published in reasonably close succession that made use of these data, by Solon (1992), Zimmerman (1992), and Altonji and Dunn (1991). The papers by Solon (1992) and Zimmerman (1992), which appeared prominently in the same issue of the *American Economic Review*, made two major contributions.[71] First, they pointed out some of the statistical problems involved in estimating the relationship between "long-run" incomes of members of the same family. Most earlier studies used single-year measures of permanent earnings and were based on nonrepresentative, homogeneous samples. Their analyses suggested that the estimates of intergenerational correlations in previous studies most likely were considerably downward biased. By using multiple years of fathers' earnings, this downward bias could be reduced. Solon also presented an estimator that most likely overestimates the correlation and thus produced a range within which the true correlation must lie. Second, their results suggested Betas between fathers' and sons' long-run incomes as high as 0.4 or 0.5, numbers that are much larger than those in the previous studies surveyed by Becker and Tomes (1986). Solon and Zimmerman obtained similar results using two different data sets, which lends additional credibility to their findings.

Solon (1992) also estimated the correlation by use of an instrumental variables (IV) method, arguing that this produces an *upward* inconsistent estimate of Beta. The argument, in brief, is to treat parental education as an omitted variable, but also to use parental

[70] Fertig (2003) shows evidence based on transition matrices for multiple cohorts of offspring but does not report the full transition matrices.

[71] Altonji and Dunn (1991) received far less attention, in part because it was published in a less well-known journal.

education as an instrument, so it is an invalid instrument. If the true direct effect of parental education is positive and it is positively correlated with parental income, such an IV estimate produces an overestimate of the intergenerational income elasticity. Thus, the OLS estimator using time-averaged parental income underestimates and the IV estimator overestimates the elasticity, bounding the parameter from below and above. Moreover, in pointing to the possibility of using an IV estimator, Solon (1992) also opened the door to estimating elasticities, reliably as it turned out, in cases where actual father–son pairs are unavailable.[72]

Mazumder (2005b), using earnings information from the U.S. Social Security Administration (SSA), examined intergenerational earnings Betas for U.S. sons and daughters with respect to fathers' earnings. His focus is on variations in the number of years across which fathers' earnings are averaged, along with several other measurement issues such as whether or not to require fathers to have positive earnings in all years, if zero earnings due to noncoverage of Social Security by registers are imputed or not, as well as whether or not zero earning offspring are included in the analysis. The results demonstrate, among other things, that attenuation from transitory variation in earnings remains substantial even after averaging fathers' earnings over up to 16 years, especially if transitory errors are characterized by autocorrelated errors. In a striking demonstration of the the impact of averaging across multiple years of parental income, Mazumder (2005b, Table 4) reported an elasticity of 0.253 (se 0.043) when only 2 years (1984–1985) of fathers' incomes are used, increasing to 0.553 (se 0.099) and 0.613 (se 0.096) when averages across 1976–1985 and 1970–1985 are used, instead. The U.S. estimates he reported thus encompass the majority of the estimates reported in Figure 10.13, excluding only Peru at the top end and Canada, Finland, Norway, and Denmark at the low end. Note, however, that in extending the number of periods over which fathers' income is averaged conflates two types of effects, namely transitory errors (whose variance is reduced) and life cycle effects (which become averaged). In the absence of estimates of λ_{Ps} and θ_{Ps}, it is hard to tell which of these is empirically responsible for the change in the elasticity.

Dahl and DeLeire (2008), also using data from the SSA but with data on noncovered years as well and using even longer time spans for fathers' earnings than Mazumder (2005b), estimated Betas for father–son and father–daughters pairs. The father–son estimates vary between 0.259 and 0.632, spanning, again, much of the observed range in the cross-country evidence on display in Figure 10.13. The father–daughter Betas range from −0.041 (which is not significantly different from zero) to 0.269. Naga (2002) used

[72] The first example of such work is that by Björklund and Jäntti (1997), which used an IV estimator based on two independent samples to estimate the intergenerational elasticity in Sweden and constructed a similar estimate for the United States. Betas estimated using actual Swedish father–son pairs are almost exactly the same as the two-sample IV estimate. There have been a large number of studies subsequently that use two-sample IV estimators including for Italy (Checchi et al., 1999), Brazil (Dunn, 2007), Australia (Leigh, 2007), and France (Lefranc et al., 2009).

father–son pairs observed at the same point in the life cycle and estimated elasticities using three methods—OLS on time-averaged data, IV, and a MIMIC latent variable estimator—and found elasticities that range from 0.297 to 0.7.

Chadwick and Solon (2002), Minicozzi (2002), and Fertig (2003) also examined Betas for women. Chadwick and Solon (2002) highlighted the importance of using family income when comparing Betas for men and women (in which case they are quite similar; when using individual earnings, Betas for women tend to be much lower). The sensitivity of Betas to sample selection rules was examined by Couch and Lillard (1998) and Minicozzi (2003). In both of those papers, sample selection issues were found to be very important for the Betas. Hertz (2005) examined racial differences in the elasticity. Estimates of Beta for the United States can also often be found in research that is either comparative or primarily about other mobility in other countries, Examples include the studies of Germany by Couch and Dunn (1997), Australia by Leigh (2007), Sweden by Björklund and Jäntti (1997), and Singapore by Ng et al. (2009).

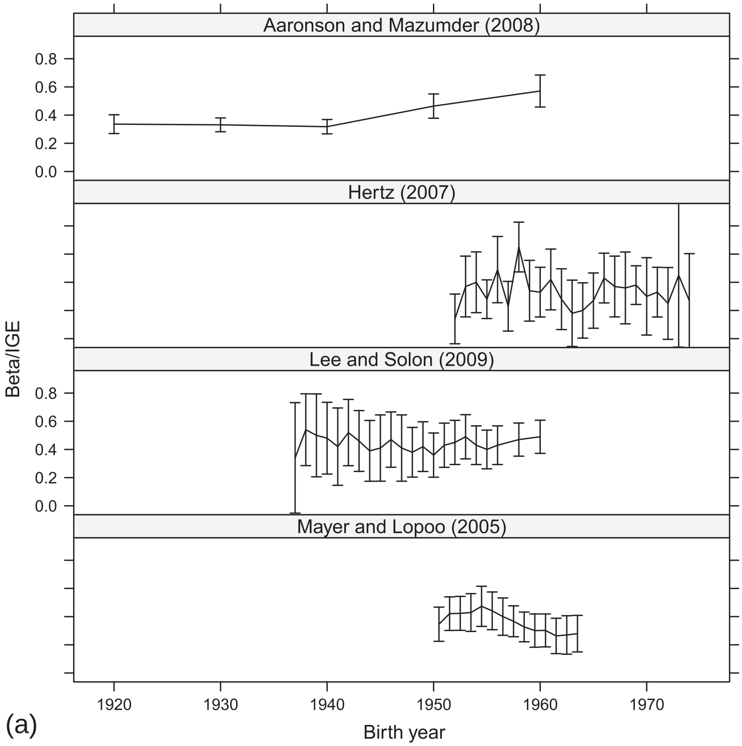
Two U.S. papers that drew attention early on to the possibility that estimates of intergenerational persistence may be subject to not only attenuation inconsistency from transitory errors in fathers' earnings or income, but also to life cycle effects in the offspring generation, were Buron (1994) and Reville (1995). Instead of adjusting for the average life cycle effects, Buron (1994) allowed earnings profiles to vary across demographic groups, which leads to a higher estimated persistence than when using the same adjustment. Reville (1995) in turn investigated how varying the age and outcome year of sons changes the estimated persistence. For instance, by following the same cohort of offspring as they age from 26–30 to 34–38 and using a 4-year average of their earnings (keeping father's earnings constant), the Pearson correlation r increases from 0.296 to 0.423 (Reville, 1995, Table 5). Hertz (2007), Lee and Solon (2009), Gouskova et al. (2010), and Chau (2012) all try to take into account biases from both transitory errors and life cycle effects.

Gouskova et al. (2010), applying the insights of both Haider and Solon (2006) and Grawe (2006), estimated earnings elasticities for father–son pairs using data from the PSID, where the fathers and sons are of the same age. Using age ranges 25–34, 35–44, and 45–54, regressing a 3-year average of sons' earnings on a 5-year average of fathers' earnings, they found elasticities of 0.29, 0.41, and 0.42, respectively. These estimates, especially the low value for the 25–34 age range, are consistent with the patterns for λ in Haider and Solon (2006). Another recent study considering the implications of the results in Haider and Solon (2006) and Chau (2012) models the income processes of both fathers and sons using heterogeneous growth profiles and autocorrelated errors. Intergenerational elasticities are then estimated based on data simulated using the parameter estimates. The U.S. estimates, based on PSID data, show an estimate of Beta of 0.392, but elasticities are as high as 0.662 when the earnings processes of sons and fathers are allowed to be different.

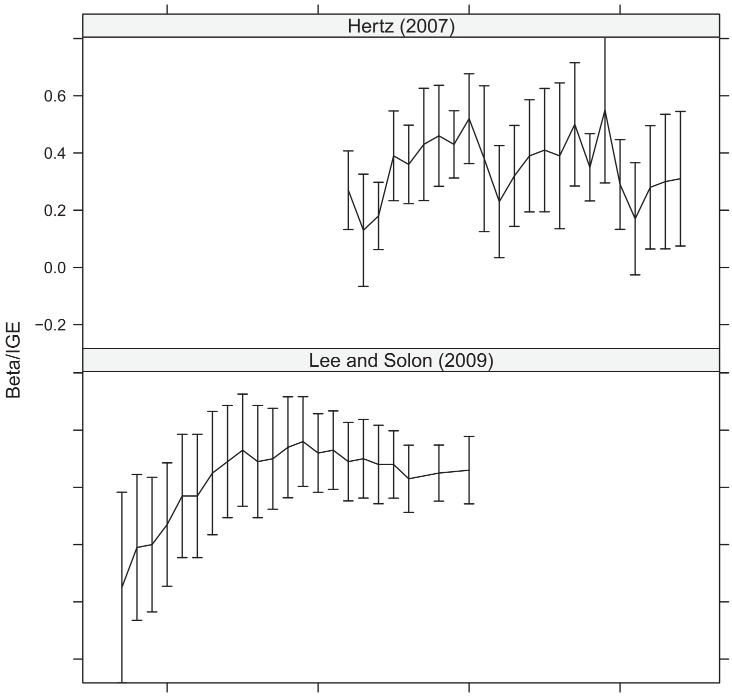
Muller (2010) tackled another complication with estimating the measurement of permanent income, namely if the elasticity varies because of shocks to parental income that take place when the offspring was living in the parental home. Parental income earned in childhood years is associated with much higher elasticities than either before the child was born or after he had left home, a result that is broadly robust with respect to standardizing the stage of the life cycle at which incomes are measured in the two generations. The results are consistent with the view that transitory shocks in childhood do affect offspring income. Although the purpose of the literature on intergenerational mobility reviewed here is not to uncover causal effects of income, this finding lends weight to the view, discussed in Section 10.5.1, that income risk may also be intergenerationally correlated.

Trends over time in intergenerational mobility in the United States, as measured by changes in Beta, have been estimated by Hertz (2007), Mayer and Lopoo (2005), Lee and Solon (2009), and, using two-sample methods, by Aaronson and Mazumder (2008). We show a selection of estimates in Figure 10.14, indexed by the birth year of the offspring, ranging from men born in the 1920s to men and women born in the early 1970s. The elasticities are evaluated at somewhat different ages, but the picture that emerges is one that suggests little systematic trend among men, with the possible exception that persistence may have increased among men from the 1940s to 1960s, mainly on display in the Aaronson and Mazumder (2008) estimates and weakly supported by both Hertz (2007) and Lee and Solon (2009). The estimates for women in Hertz (2007) and Lee and Solon (2009) suggest increasing persistence for the early cohorts but little change from around 1960 onward. The differences across studies suggest care must be taken in interpreting trends based on but a few data points and sets of definitions. The large confidence intervals around each point estimate also highlight the importance of statistical inference. Indeed, all the confidence intervals in the series from Hertz (2007), Lee and Solon (2009), and Mayer and Lopoo (2005) overlap. Although this does not mean there cannot be significant differences between point estimates, it does warrant some caution.

The IGE (Beta) is related to a "global" log-linear regression, forcing the slope of the conditional expectation of offspring log income to be a linear function of parent log income. There are many ways to relax the assumption that the slope is the same everywhere. Differences in the slope at different levels of parental income can be motivated by theoretical concerns. A commonly cited concern is the potential presence of borrowing constraints with respect to parental investments in child human capital (Becker and Tomes, 1986; Bratsberg et al., 2007; Grawe, 2004b). Bratsberg et al. (2007) fit a polynomial in parental income to the data for the United States drawn from the NLSY to allow for a flexible shape between offspring and parental income. They found that a second-order polynomial in parental income provides a reasonable fit for U.S. data. The IGE based on a log-log regression is 0.542, whereas those based on the polynomial imply elasticities of 0.489, 0.575, and 0.646 at the 10th, 50th, and 90th percentiles of parental income, respectively. Couch and Lillard (2004) demonstrated that these results are highly sensitive to the procedure applied. Using both second- and third-order



(a)



(b)

Figure 10.14 (See legend on next page) polynomials in both the log and the level of parental income, they estimated elasticities in the first, third, and fifth quintile groups of fathers' income to be 0.124, 0.234, and 0.292 using a quadratic, and 0.219, 0.230, and 0.171 for the cubic polynomial, compared to 0.158 in the log-log. Thus, using the second-order polynomial, elasticities increase monotonically across fathers' income but, using the third-order polynomial, they increase to decline at higher levels. Another option is to estimate the conditional mean (and, by implication, its slope) nonparametrically, for instance using kernel regression.

The elasticity is a measure of average *persistence* of income rather than of *mobility*. In other words, the regression coefficient on father's log (permanent) earnings tells us how closely related, on average, an offspring's economic status is to that of his or her parent. It is quite possible for two distributions to have highly similar average persistence, but for one to have substantially more mobility around that average persistence. The elasticity can thus be the same, but arguably the distribution with a greater residual variation—variability around the average persistence—is the one with greater mobility (see the discussion of the Gottschalk and Moffitt "BPEA" measure in Section 10.3). Moreover, two distributions with the same regression slope may have quite different, and varying, conditional variances around that slope. For instance, a distribution with a "bulge" in the variance at low levels of fathers' earnings, that is, a pear-shaped bivariate distribution, will exhibit relatively more mobility at the low end of the distribution than will a distribution with a constant conditional variance.

One approach is to examine both the regression coefficients and residual variances. Other approaches, such as nonparametric bivariate density estimates, similar to Figure 10.4 in the intragenerational case in Section 10.3, would in principle be available (see, e.g., Bowles and Gintis, 2002). Very few studies take that route, however. Quantile regression (Koenker, 2005) can also be used to examine the conditional distribution of offspring income, conditioned on parental income. Although the slopes of the conditional quantiles of offspring income can be of interest in and of themselves, we tend to find what they say about the full conditional distribution of greater interest than the slopes of individual quantiles (cf. the discussion of this in Section 10.3.3). In the prototypical homoscedastic regression, where the variance (or indeed, any higher moments) of the

Figure 10.14 Trends in U.S. intergenerational income persistence. *Note: The estimates in Lee and Solon (2009) are the elasticities for different outcome years at age 40, presented here by subtracting 40 from the outcome year and are derived using a 3-year average of parental income. Mayer and Lopoo (2005) estimate elasticities for 4-year birth cohorts, which are centered here, observe offspring at age 30, and use a 7-year average of parental income (at ages 19–25). Hertz (2007) presents elasticities at age 25 and uses a 3-year average of income. His estimates further control for panel attrition. Aaronson and Mazumder (2008) uses two-sample methods applied to (IPUMS) census data, with elasticities applying to 35- to 44-year-olds, here centered at age 40. Source: Aaronson and Mazumder (2008, Table 1, column 6), Hertz (2007, Table 4), Mayer and Lopoo (2005, Table A1), and Lee and Solon (2009, Table 1).*

residual does not depend on the explanatory variable, the quantile regression slopes should all be straight lines with slopes equal to the conditional mean and median. Deviations from these patterns are informative of variations in the shape of the conditional distribution.

Eide and Showalter (1999) estimated quantile regressions for several percentiles using PSID data on father–son pairs where the sons are 25–34 years old, using a 3-year average of parental income and 7-year average of sons' earnings. They found a Beta of 0.34 and slopes of the conditional quantiles with respect to parental income of 0.77 at the 5th percentile, 0.47 at the 10th percentile, 0.37 at the 50th percentile (median), 0.17 at the 90th percentile, and 0.19 at the 95th percentile. That is, they (mostly) find the slope to be decreasing in the percentile but also that the Beta is lower than the slopes of the quantiles up to the 75th percentile.[73]

Conditional quantiles can be combined with nonparametric techniques to allow for the slope to change flexibly. We illustrate this in Figure 10.15 from Lee et al. (2009), who used PSID data for U.S. sons and fathers to nonparametrically estimate the conditional quantiles of sons' income conditioned on fathers. We can see that the slopes of lower quantiles tend to be steeper at low parental income than for the higher quantiles and that the slopes tend to level of as parental income increases.

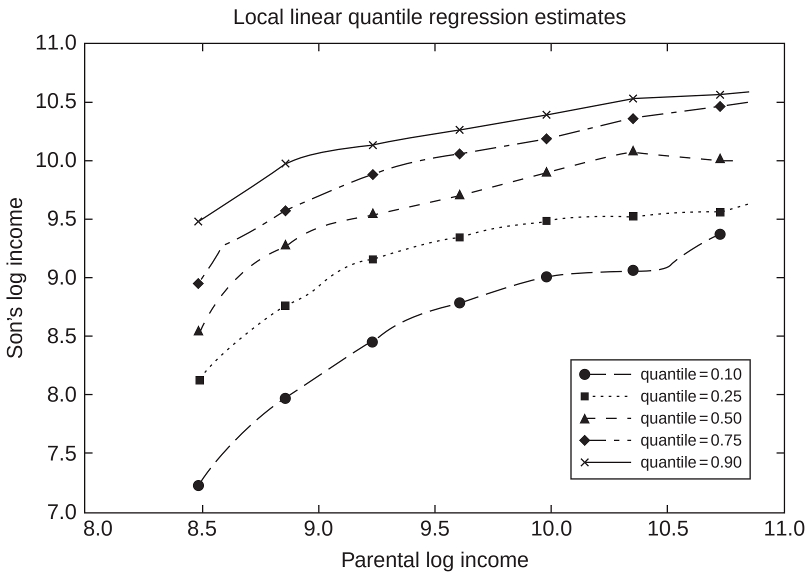


Figure 10.15 Intergenerational income persistence: nonparametric quantile regression for U.S. father–son pairs. *Note: Estimates based on PSID father–son pairs as prepared by Minicozzi (2003). Sons' income is the average of labor income at ages 28 and 29, and parental income is predicted parental income as defined by Minicozzi (2003). Source: Lee et al. (2009, Figure 1).*

[73] See also Grawe (2004a) for additional U.S. estimates.

Asymmetries in intergenerational mobility can be straightforwardly described using transition matrices, a simple but underused device for illustrating intergenerational mobility. In allowing for fairly general patterns of mobility, mobility or transition matrices offer the additional advantage of allowing for asymmetric patterns, for example more mobility at the top than at the bottom. To illustrate, we show in panel A of Table 10.5, a decile transition matrix for U.S. fathers and sons.

The cell entries show, for each decile group of origin (i.e., fathers' decile group), the percentage of sons in each destination decile group. Specific aspects of the transition matrix tend to be highlighted. For instance, the main diagonal shows the percentage of sons who remain in their father's decile group. One descriptive statistic is the sum of the main diagonal probabilities (the matrix trace), in this case 165. With ten income classes, there is origin independence if each entry in the table is 10%, which implies an average "excess" immobility relative to origin independence of 6.5% points. Conversely, 83.5% of U.S. sons are in a different decile group than their fathers. The Normalized Trace index (Shorrocks, 1978b) for this matrix is $(10-165/100)/(10-1)=0.93$.

The corner probabilities are often of special interest also. In this case, 22% of the sons of the poorest 10th of fathers are in the poorest 10th themselves, whereas 26% of the sons of the richest 10th of fathers are in the richest 10th. Conversely, upward mobility from the lowest 10th is $100-22=78\%$, and downward mobility from the highest 10th is $100-26=74\%$. By contrast, 7% of sons of poorest fathers and 3% of the richest end up in the top and bottom decile groups, respectively. Somewhat to our surprise, we are unable to illustrate an application of dominance analysis to examine the change across cohorts in U.S. intergenerational mobility. We are unaware of a comparable transition matrix for a later or earlier cohort.

A final observation can be made regarding the "shape" of the transition matrix. Transition matrices for bivariate normal data, such as the simulated data in O'Neill et al. (2007) or the illustrations of the consequences of different r in Björklund and Jäntti (1997), are symmetric. For instance, the two corners on the main diagonal are equal as are the corners on the antidiagonal, and the upper triangle is the mirror of the lower triangle. The U.S. father–son transition matrix clearly exhibits very little symmetry of this sort. The lack of symmetry implies that both mobility and persistence may be different across the distribution, and of course that the data are unlikely to be well described by a bivariate log normal distribution.

Recently, Bhattacharya and Mazumder (2011) proposed a set of measures based on the bivariate percentile distribution, focusing specifically on upward and downward mobility relative to a parameter τ that specifies the number of percentiles one needs to move up to be considered upward mobile, illustrating their approach by comparing mobility differences between racial groups in the United States using data on men from the NLSY. Whites are found to be distinctly more likely to move upward than blacks.

Table 10.5 Intergenerational decile transition matrices for earnings, father–son pairs, Canada and the United States

A. United States

Father	Son 1	2	3	4	5	6	7	8	9	10
1	22	18	10	10	11	11	5	5	2	7
2	9	15	16	15	9	9	9	5	9	5
3	9	10	12	17	15	9	9	7	7	5
4	17	9	10	12	3	15	9	11	7	7
5	12	7	12	6	14	9	12	10	12	8
6	7	11	6	10	11	13	13	11	7	11
7	8	7	12	9	11	9	16	13	9	5
8	8	8	8	11	10	7	11	15	13	8
9	4	8	8	5	9	11	7	9	20	19
10	3	8	6	7	7	5	10	16	11	26

B. Canada

Father	Son 1	2	3	4	5	6	7	8	9	10
1	16	14	12	11	10	9	8	7	7	7
2	13	13	12	12	11	10	9	8	7	6
3	11	11	12	12	12	11	10	8	8	7
4	10	10	11	11	11	11	11	10	8	7
5	9	10	10	10	11	10	11	11	10	8
6	9	9	10	10	10	11	11	11	10	9
7	8	9	9	9	10	10	11	11	11	11
8	8	8	8	9	9	10	11	12	12	12
9	8	8	8	8	8	10	10	12	13	15
10	8	8	8	8	8	9	10	11	13	18

Note: The cell entries show, for each decile group origin (referring to fathers), the percentage of sons in each destination decile group. U.S. estimates are based on SIPP matched to Social Security earnings. Fathers' earnings are averaged across 1979–1985 and sons' across 1995–1998. Canadian data are based on tax records. Fathers' earnings are averaged across 1978–1982 and sons' earnings across 1993–1995.

Source: Mazumder (2005a, Table 2.2) and Corak and Heisz (1999, Table 6).

10.5.3 Cross-National Comparative Evidence on Intergenerational Associations

We now turn to examining evidence on intergenerational income mobility in other (mostly rich) countries. To illustrate the importance of how mobility is measured for cross-country rankings, we start this subsection by reporting results from two recent papers, each of which compares three countries. Corak et al. (2013) compared earnings mobility between fathers and sons in Canada, Sweden, and the United States. Their focus is on comparing upward and downward mobility, but we rely here on their three estimates of persistence: Beta (IGE), the Pearson correlation *r*, and the Spearman rank correlation are reported in Table 10.6 along with the ranking of the three countries in each case. The estimated Betas are in line with those found in previous research and show intergenerational income persistence to be the greatest in the United States, followed by Canada and Sweden. The ranking by the product–moment correlation *r* is the same, but now the U.S. point estimate is much closer to those of Canada and Sweden. By contrast, according to the rank correlations, Canada has the lowest persistence, and Sweden and the United States are tied. This, arguably the preferred scalar index of persistence (as it most clearly abstracts from differences in marginal distributions), suggests a very different ordering of countries with respect to intergenerational mobility than that on display in the "Great Gatsby" curve of Figure 10.13.

Eberharter (2013) estimated persistence in terms of Betas for disposable income among men and women in Germany, the UK, and the United States, using data from the U.S. PSID, the German SOEP, and the UK BHPS. The elasticity estimates are reported in the left panel of Figure 10.16 together with the 95% confidence intervals. This is a rare study because it presents estimates for several countries using measures of disposable income. It is also unusual to pool sons and daughters, although that choice is arguably well motivated when the purpose is to examine the persistence in living standards.

Table 10.6 Intergenerational earnings mobility in Canada, Sweden and the United States: Beta, *r*, and the rank correlation

	Beta		r		Rank correlation	
Country	**Estimate**	**Rank**	**Estimate**	**Rank**	**Estimate**	**Rank**
Canada	0.26	(2)	0.23	(2)	0.24	(1)
Sweden	0.25	(1)	0.21	(1)	0.30	(2)
United States	0.40	(3)	0.26	(3)	0.30	(2)

Note: Canadian estimates rely on tax records. Father's earnings are a 5-year average and son's a 3-year average 1997–1999 when they were 31–36 years old. Swedish estimates, also based on tax records for earnings, rely for fathers on 20 years of earnings data measured at ages 30–60 and for sons on an 11-year average across ages 30–40. The U.S. estimates stem from the Survey of Income and Program Participation panels using earnings from Social Security records. Fathers' earnings are a 9-year average between 1979 and 1986 when they were 30–60 years old. Sons' earnings are a 5-year average between 2003 and 2007 in years they were at least 28 years old.
Source: Corak et al. (2013, pp. 10–11).

Although Eberharter (2013) did not report rank correlations, these results bring out quite forcefully the importance of being wary of changes in marginal distributions across the cohorts, especially when comparing estimates from different countries.[74] As can be seen by comparing the left panel of Figure 10.16, which plots the elasticities, with the right panel, which reports the implied (Pearson) correlations r, the results are dramatically different in the two cases. The United States has a substantially higher elasticity than either Germany or the UK (0.68 as opposed to 0.48 and 0.50), but when we derive

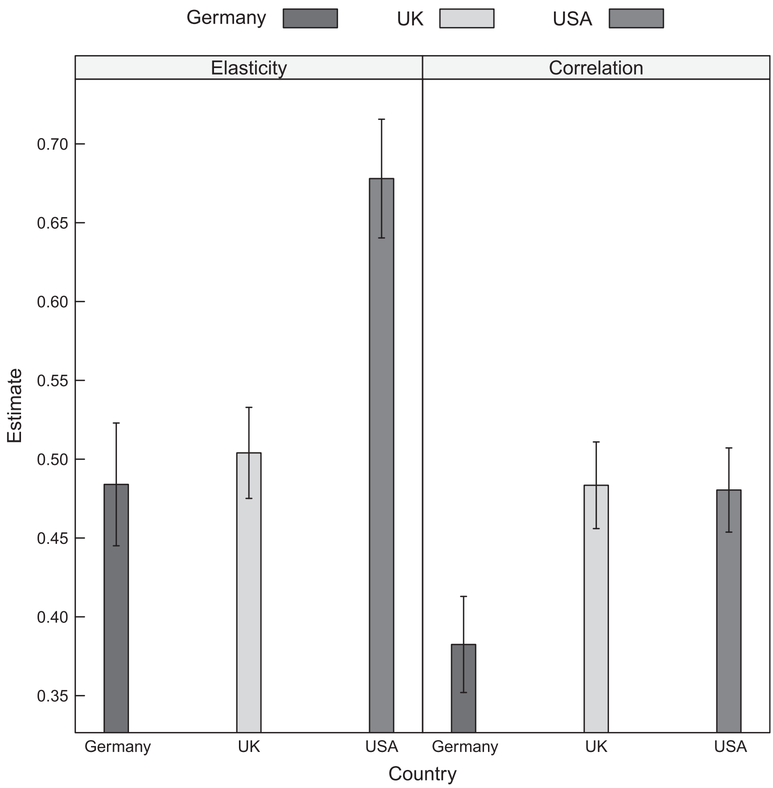


Figure 10.16 Intergenerational persistence of disposable income: elasticities versus correlations. *Note: Error bars show 95% confidence intervals. Estimates are for posttax, posttransfer income for all individuals (for sons and daughters combined). Offspring incomes are observed for those older than 24 who are out of full-time education and are averaged across 2005–2009 (Germany), 2003–2007 (USA), and 2004–2008 (UK). Parental income are observed as offspring were 14–20 years old and are averaged across 1988–1992 (Germany), 1987–1991 (USA), and 1991–1995 (UK). Eberharter (2013) reports standard deviations in parental and offspring generations for full samples rather than the estimation samples, so the estimated implied correlations, obtained using* $\rho = \sigma_P/\sigma_O\beta$ *are approximate only. Source: Authors' elaborations based on Eberharter (2013, Tables 1 and 2).*

[74] See, for example, Björklund and Jäntti (2009).

the correlations, the UK has a correlation that is higher than that in the United States, and Germany's is substantially lower than either of those.[75] It is not possible, of course, to infer what the rank correlations are from the Betas and *r*.

Thus, even confining ourselves to scalar measures of mobility, switching between Beta and the two correlations leads to rank reversals. The fact that Sweden and the United States, two countries that inhabit very different regions in the "Great Gatsby" curve diagram, have equal mobility as measured by the rank correlation, is particularly notable.

Most studies of intergenerational income persistence and mobility were inspired by the U.S. studies of Solon (1992), Zimmerman (1992), and Altonji and Dunn (1991). An exception is the study of intergenerational mobility in the UK (Atkinson, 1981b; Atkinson et al., 1983; cited several times by Solon, 1989, 1992, possibly serving as inspiration for the U.S. studies). Intergenerational income persistence in the UK, especially the question of whether it has changed, has been subject to substantial controversy recently. We therefore start our discussion of single-country studies with UK evidence.

The early estimates of father–son Betas in Atkinson (1981b) and Atkinson et al. (1983) using a geographically limited and truncated sample were around 0.44. Atkinson et al. (1983, p. 111) discussed the impact of measurement error in parental income, finding that for plausible value of the signal-to-noise ratio, the true Beta might well be at least 0.5. Dearden et al. (1997), using data from the cohort study of 1958-born children (the NCDS), estimated Beta to be around between 0.29 (OLS) and 0.58 (2SLS). Later studies by Blanden and Machin (2008), Blanden et al. (2010), and Blanden et al. (2013) have generated a reasonably wide range of UK estimates.

One particularly contested UK finding is that mobility has decreased, based on the finding that the IGE estimated for the cohort born in 1958 (NCDS) is greater than the IGE for the cohort born in 1970 (BCS). Depending on estimation method, the elasticity increased from 0.31 to 0.33 (OLS) or 0.33 to 0.50 (2SLS), both measured for sons at age 34 (Blanden and Machin, 2007). Most recently, using a single-year measure of parental income and no controls for parental age, Blanden et al. (2013) reported an increase in the IGE between NCDS and BCS cohorts from 0.211 to 0.278 for parent–son pairs, corresponding to a difference of 0.067 (se 0.034). These estimates have been widely referred to in UK public policy debates about social mobility, as discussed recently by Goldthorpe (2013).

The UK debate provides several lessons. First, two estimates provide little evidence about the existence of a trend. U.S. estimates for different birth cohorts vary quite substantially; see Figure 10.14, where there is no apparent trend. Moreover, different data

[75] The orderings are statistically robust: the confidence interval for the U.S. elasticity does not overlap those for Germany or the UK and, as the intervals for the correlations between Germany on the one hand and the United States and the UK do not overlap, pairwise *t*-tests reject the null that the correlations are the same.

sources and estimation methods may generate different results. For example, Nicoletti and Ermisch (2007) derived Betas for Britain using two-sample methods applied to BHPS data. They estimated relatively stable elasticities and correlations for cohorts born between 1950 and 1960. For the cohorts born between 1961 and 1972, elasticities rose somewhat over time but correlations are stable. These results are only partially consistent with the estimates derived from the BCS and NCDS cohorts. Second, data quality has serious implications for public policy. Part of the UK controversy centers around whether the two cohort studies in question, the NCDS and BCS, have sufficiently comparable data. Third, measures of intergenerational *income* mobility may change over time in a different way from measures relating to other concepts of intergenerational economic and social mobility. These differences may in turn be informative of the nature of societal change.

The possibility that intergenerational income persistence in the UK has increased, but class mobility has not, led Erikson and Goldthorpe (2010) to examine mobility in the earnings/income and class spaces. They concluded that problems with the measurement of income in the parental generation render the finding of an increase in income persistence suspect, and they emphasized the stability of social class mobility over time as indicating that there has been little change in intergenerational mobility in the UK.

More recently, Blanden et al. (2013) used an approach proposed by Björklund and Jäntti (2000) to decompose the *r* (strictly speaking, the partial correlation) into the correlation of "class-predicted" incomes, the correlation of deviations of actual from class-predicted incomes, and their cross-correlations. Their results are consistent with there being stable class mobility, as suggested by there being no contribution (but in fact, a small negative one) from the "class-predicted" income correlation to the change, whereas all three correlations involving the residuals contributed to an increased partial correlation. The results can be interpreted as saying income and class mobility decreasingly capture the same phenomena, as the relationship between income and class appears to be different in the later than in the earlier cohort. The discussion of these results by Blanden et al. (2013), Erikson and Goldthorpe (2010), and Goldthorpe (2013) provides valuable insights into the scientific and public debates about social and economic mobility.

A key conclusion that we draw about the UK debate, not least in light of the divergent U.S. estimates of both levels and trends, is that much richer data than those provided by the NCDS and BCS cohort studies are needed to draw firm conclusions about the level and trend in UK income mobility.[76] It is also possible that class and income mobility are diverging because the processes that generate transitory errors are changing in ways that

[76] We note, in passing, that the dominance analysis conducted by us of the income quintile group transition matrices reported by Blanden et al. (2010, Table 3) for the NCDS and BCS suggest that, except for the cells (5,3) and (5,4) all BCS–NCDS differences in the cumulated matrices are positive (but those entries are negative). Thus, there is no dominance between the two cohorts.

suggest intergenerational advantage is increasingly transmitted through deviations from the systematic components of income. In our view, the UK debate underlines the need for high-quality data to resolve what has turned out to be a question of great social concern.

Corak and Heisz (1999) provided Betas for both earnings and total market income for Canadian father–son pairs, using (at most) a 5-year average of parental income and a single year for sons' in 1995 at which point they are 29–32 years old. They find elasticities for earnings of 0.131 and for market income of 0.194. In addition to transition matrices, discussed later, they also estimated the conditional expectation, and its slope, of sons' earnings with respect to fathers', nonparametrically. They found that the elasticity varies substantially and quite nonmonotonically across the distribution of fathers' earnings.

Leigh (2007) estimated the intergenerational earnings elasticity for Australian men using two-sample methods. For men born in 1949–1979, he estimated an elasticity of 0.181. This compares to a U.S. elasticity for a similar cohort of sons, obtained using similar estimation methods, of 0.325. The difference is statistically insignificant, but still suggests Australian persistence is lower. His results for older cohorts vary substantially, however. For men born in 1911–1940 and 1919–1943, the point estimates are 0.26, but for men born in 1933–1962, the estimate is 0.413. Gibbons (2010) estimated intergenerational mobility for New Zealand father–son and father–daughter pairs of 0.25 and 0.17, respectively.

Lefranc (2011) used two-sample methods to estimate Betas for cohorts of men born between 1931 and 1975 in France. The estimates, which start at 0.626 for men born 1931–1935 decline to 0.441 for cohorts born 1956–1960 and increase thereafter, being 0.559 for cohorts born 1971–1975. Estimates for Spain are provided, e.g., by Cervini-Plá (2009) and for Italy by Mocetti (2007), both of which are high by international standards at about 0.4 and 0.5, respectively.

Pekkala and Lucas (2007) estimated intergenerational elasticities for Finnish cohorts born between 1930 and 1970, using census data on annual earnings for offspring and family income for parents. The intergenerational elasticities declined substantially; for sons from more than 0.30 to around 0.20, and for daughters from 0.25 to around 0.15 for cohorts born in 1930 to those born in 1950 and later. It may be of special interest to note that Pekkarinen et al. (2009) found comprehensive school reform, treated as a quasi-experiment, reduced the Finnish Beta by almost a third. The Norwegian trend studies have focused on the post-1950 cohorts. Bratberg et al. (2007) found a small decline in father–son and father–daughter elasticities from 1950 to 1965 cohorts. However, Hansen (2010) reported that this result does not hold when using the income of both parents. Instead, she found a small increase in the elasticities for the 1955–1970 cohorts. This difference suggests an increasing role for mothers, which has not been much explored in the literature. The Beta for Swedish father–son pairs is around 0.25 (see, e.g., Björklund and

Chadwick, 2003), but much higher at the top of the distribution (Björklund et al., 2012). Estimates from Denmark suggest quite low levels of persistence (e.g., Bonke et al., 2005).

Lefranc et al. (2013) estimated Betas for Japanese sons and daughters using two-step sample methods. The estimates for men are all quite close to 0.35. For daughters, estimates vary between 0.182 and 0.367. The evidence on whether or not the Betas increased for younger cohorts is mixed, at best. Ueda (2009) used instrumental-variable techniques to estimate elasticities for men and women in Japan also and found elasticities of around 0.411–0.458 for men, and 0.229–0.361 for women, depending on marital status and the use of family or individual income.

Nonlinearities in the parent–child conditional income expectation were explored in a multicountry study by Bratsberg et al. (2007), who found the data for the United States, UK, Denmark, Norway, and Finland all suggested the relationship is convex, with elasticities low at low levels of parental income, and increasing thereafter. At all quantiles of parental income, the elasticities are lower for the Nordic countries than for the UK and the United States. Interpreted in terms of borrowing constraints on investments in child human capital, the results suggest capital market imperfections may be more of an issue not at the bottom but more around the middle of the distribution of parental income.

Raaum et al. (2007) tackled another question in a multicountry study, namely how the mobility of daughters compares with that of sons across countries. Drawing on Chadwick and Solon (2002) and Björklund and Chadwick (2003), they found women's intergenerational income persistence is very similar across countries relying only on individual earnings. When family earnings are used for both men and women, the country ordering of intergenerational persistence for women looks very much like that for men. Using a framework that involves the intergenerational transmission of human capital endowments, assortative mating, and labor supply that responds to both own and spouse's wage, they inferred that female labor supply is likely more (negatively) responsive to husband's earnings in the UK and especially the United States than in the Nordic countries.

We proceed to compare transition matrices across countries. To illustrate, consider the decile group transition matrices for the United States and Canada shown in Table 10.5 and derived from Mazumder (2005a) and Corak and Heisz (1999). Using the dominance approach discussed in Section 10.3.2, we can cumulate the transition matrices and take the United States–Canada difference. This leads to the results shown in Table 10.7. The vast majority of the cell entries are positive, suggesting Canada dominates the United States. However, given the two negative entries in cells (10,1) and (10,9), this result does not hold, strictly speaking.[77]

[77] We have not forced the rows or columns of either transition matrix to sum to 1, as they should in a bistochastic matrix. The U.S. matrix in particular fails this condition (most likely due to a smaller sample size and rounding error). If we do force the rows to sum to 1, the negative entries vanish, and we have dominance.

Table 10.7 Cumulated differences in intergenerational mobility tables across earnings decile groups for father–son pairs in Canada and the United States (USA–CAN)

	Son									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Father										
1	6	10	9	8	9	11	8	6	1	1
2	2	9	11	13	12	14	11	6	3	2
3	1	6	8	16	18	18	15	8	4	2
4	8	11	13	21	16	20	15	10	4	2
5	10	12	15	19	17	19	15	9	7	4
6	9	12	11	15	14	19	17	11	5	4
7	8	9	12	15	15	18	22	18	10	3
8	8	9	11	17	17	17	21	21	13	2
9	4	5	7	9	10	12	12	9	9	2
10	−1	0	0	2	2	0	0	2	−1	0

Note: Cell entries are in percent. See notes to Table 10.5.
Source: Authors' derivations using transition matrices shown in Table 5 from Mazumder (2005a) and Corak and Heisz (1999).

Recall from Figure 10.16 that between the Betas and *r*s for disposable income, the United States and the UK were reranked, whereas Germany was least persistent in both. In Table 10.8, we illustrate again the use of the dominance approach, this time using quintile group transition matrices also from Eberharter (2013). The differences in the cumulated transition matrices suggest that Germany dominates both the UK and United States (all entries in the United States–Germany and UK–Germany matrix are positive), but that the United States and UK cannot be ordered. Note, however, that there is only one strictly positive entry in cell (3,2), indicating the United States is close to dominating the UK.

10.5.4 Evidence on Sibling Correlations

In this section, we show evidence on sibling correlations and relate them to intergenerational correlations. Why are sibling correlations of interest in the study of intergenerational income mobility? One way to motivate the interest in intergenerational mobility is to argue that it is related to equality of opportunity (see Section 10.2). A society in which a person's position is heavily dependent on the family he/she is born into is one in which there is likely to be less equality of opportunity than one in which intergenerational persistence is very low.[78] But if we would like to understand how important family background is for the distribution of economic status, a focus on parent–child association captures only one part of the association. A fuller (but still incomplete) accounting of the

[78] As we argued earlier, and to underline a point made repeatedly in the literature, the link between intergenerational mobility and equality of opportunity is far from straightforward.

Table 10.8 Cumulated differences in intergenerational transition matrices in disposable income among all persons for Germany, the UK, and the United States

	A. US–Germany						**B. US–UK**						**C. UK–Germany**				
	Offspring						**Offspring**						**Offspring**				
	1	**2**	**3**	**4**	**5**		**1**	**2**	**3**	**4**	**5**		**1**	**2**	**3**	**4**	**5**
Father						**Father**						**Father**					
1	3	5	5	1	0	1	−10	−1	−1	0	0	1	14	6	7	2	0
2	9	11	4	2	0	2	−11	−5	−2	−6	0	2	20	16	6	8	0
3	9	18	6	2	0	3	−11	1	−4	−9	0	3	20	18	11	11	0
4	9	18	9	9	0	4	−8	−3	−12	−10	−1	4	17	20	21	19	1
5	4	13	1	2	0	5	−10	−11	−21	−20	−1	5	15	24	22	23	1

Note: Cell entries are in percent. See notes to Figure 10.16.
Source: Authors' calculations from Eberharter (2013, Table 3).

importance of family background can be done by comparing the economic status of siblings. It turns out that the *sibling correlation* can be thought of as an R^2 of family background, capturing the importance of factors that siblings share in (most often) the variance of log income or earnings. Although part of what siblings share is parental income, a large part is not. That is why sibling correlations are useful in assessing the importance of family background in the distribution of economic status.

To clarify the interpretation of a sibling correlation, we follow the exposition of Solon et al. (1991). Suppose that we observe annual income, assumed to equal long-run income plus transitory errors, assumed to be classical. The natural logarithm of income in year t, y_{ijt}, for sibling j in family i, for brevity, assumed to be measured as deviations from the population average, is modeled as

$$y_{ijt} = a_i + b_{ij} + v_{ijt}, \tag{10.24}$$

where a_i is a permanent component common to all siblings in family i, and b_{ij} is a permanent component unique to individual j, which captures individual deviations from the family component. The error term v_{ijt} picks up deviations of annual income from long-run income. The family and individual components are orthogonal by construction, so the long-run income variance is the sum of the family and individual component variances, $\sigma_a^2 + \sigma_b^2$. The share of the variance of long-run income that can be attributed to family background is

$$\rho = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_b^2}. \tag{10.25}$$

This share coincides with the Pearson correlation in long-run income of randomly drawn pairs of brothers, which is why ρ is called a sibling correlation. As the conceptual model underlying the sibling correlation is defined in terms of variances, it can only vary between 0 and 1 (i.e., negative correlations are ruled out).

A sibling correlation can be thought of as an omnibus measure of the importance of family and community effects. It includes anything shared by siblings—parental income and parental influences such as aspirations and cultural inheritance, as well as neighborhood influences such as from school, church, and peers. Genetic traits not shared by siblings, differential treatment of siblings, time-dependent changes in neighborhoods, and so on are captured by the individual component b_{ij}. The more important the effects that brothers share, the larger is the brother correlation.

Part of what siblings share in a is parental income. A useful analytical insight is that (assuming for ease of exposition marginal distributions are in steady state) the brother correlation in income can be thought of as the sum of the intergenerational income correlation squared and the correlation of other factors siblings share but that are orthogonal to income:

$$\rho = r^2 + \text{correlation of other shared factors}. \tag{10.26}$$

When the steady-state assumption is untrue, the first part of the sum on the right-hand side of Equation (10.26) also involves the marginal distributions of income in the two generations. This decomposition allows us to apportion the overall importance of family background, as captured by ρ, onto that part accounted for by intergenerational persistence, measured by Beta or r, and the other factors siblings share that affect income.

Evidence about sibling correlations in earnings (and income) was surveyed by Solon (1999), Björklund and Jäntti (2009), and also by Schnitzlein (2014), who provided new estimates for Denmark, Germany, and the United States. We show in Table 10.9 evidence based mostly but not only on long-run earnings for several countries. The evidence is based on three main methods for estimating the variance components that constitute the sibling correlation: (unbalanced) ANOVA, restricted maximum likelihood estimates (REML), and generalized method of moments. As Björklund et al. (2009) reported, whether or not the transitory errors are allowed to be autocorrelated has a big impact on the estimated sibling correlations. Allowing errors to be autocorrelated tends to reduce the individual variance, so increasing the estimated sibling correlation, so cross-country comparisons should be made across similarly defined models.

Although there are multiple estimates for several of the countries, we have sibling correlations in earnings or income for no more than seven countries for brothers, and six for sisters. The estimates for Nordic countries are low (and by far the lowest for Norway), highest for China, and of similar magnitude in Germany and the United States. For men, 43–49% of the variance in long-run earnings in Germany and the United States is accounted for by family background. This compares to 14% in Norway and 20–25% in the other Nordic countries. The ordering is similar, but levels for women are lower across the board. Family background accounts for 30–39% of long-run earnings in Germany and the United States and between 11% and 23% in the Nordic countries.

The sibling correlation is a ratio of the variance of the family component in income to the variance of long-run income. In the spirit of the "Great Gatsby" curve, shown in Figure 10.13, it is of interest to compare now another measure of persistence, the sibling correlation, with another measure of cross-sectional inequality, namely that of permanent earnings or income. We plot in Figure 10.17 the brother and sister correlations against the standard deviation of (the natural logarithm of) permanent earnings/income for those cases listed in Table 10.9 where we have been able to find all variance components.[79] In each panel, we have drawn the least-squares regression line.

Despite the small number of countries some insights can be gained. Among men, the estimated levels of permanent income inequality are consistent with very different degrees to which family background accounts for long-run earnings. Finland, Denmark, and Norway have a standard deviation of log permanent earnings on either side of 0.4, as

[79] Figure 10.13 has a single point for each country, whereas we have included repeated observations for a country in Figure 10.17 for some cases.

Table 10.9 Sibling correlations in earnings and income

Brothers					Sisters				
Denmark	0.23	1951–1968	ANOVA	Björklund et al. (2002)	Denmark	0.19	1958–1971	REML	Schnitzlein (2013)
Denmark	0.20	1958–1971	REML	Schnitzlein (2013)	Finland	0.13	1950–1960	ANOVA	Österbacka (2001)
China	0.57	Not reported	REML	Eriksson and Zhang (2012)	Finland	0.11	1955–1965	ANOVA	Björklund et al. (2004)
Finland	0.26	1953–1965	ANOVA	Björklund et al. (2002)	Germany	0.39	1958–1971	REML	Schnitzlein (2013)
Finland	0.26	1950–1960	ANOVA	Österbacka (2001)	Sweden	0.15	1951–1968	ANOVA	Björklund et al. (2004)
Finland	0.24	1955–1965	ANOVA	Björklund et al. (2004)	Sweden	0.23	1953	REML	Björklund et al. (2010)
Germany	0.43	1958–1971	REML	Schnitzlein (2013)	Norway	0.12	1953–1969	ANOVA	Björklund et al. (2004)
Norway	0.14	1950–1970	ANOVA	Björklund et al. (2002)	USA	0.34	1947–1955	REML	Mazumder (2008)
Norway	0.14	1953–1969	ANOVA	Björklund et al. (2004)	USA	0.28	1951–1958	ANOVA	Solon et al. (1991)
Sweden	0.37	1962–1968	GMM	Björklund et al. (2009)	USA	0.29	1958–1971	REML	Schnitzlein (2013)
Sweden	0.25	1953	REML	Björklund et al. (2010)					
Sweden	0.25	1948–1965	ANOVA	Björklund et al. (2002)					
Sweden	0.22	1962–1968	REML	Björklund et al. (2007a)					
Sweden	0.19	1951–1968	ANOVA	Björklund et al. (2004)					
USA	0.49	1947–1955	REML	Mazumder (2008)					
USA	0.45	1944–1952	REML	Levine and Mazumder (2007)					
USA	0.45	1951–1958	ANOVA	Solon et al. (1991)					
USA	0.43	1951–1967	ANOVA	Björklund et al. (2002)					
USA	0.45	1958–1971	REML	Schnitzlein (2013)					

Note: Estimates are all based on multiyear averages of earnings or income, adjusted for stage in lifecycle. We have relied in part on the compilation of evidence in Schnitzlein (2013) in constructing this table.
Source: Schnitzlein (2013) and authors' compilation from sources listed in the last column.

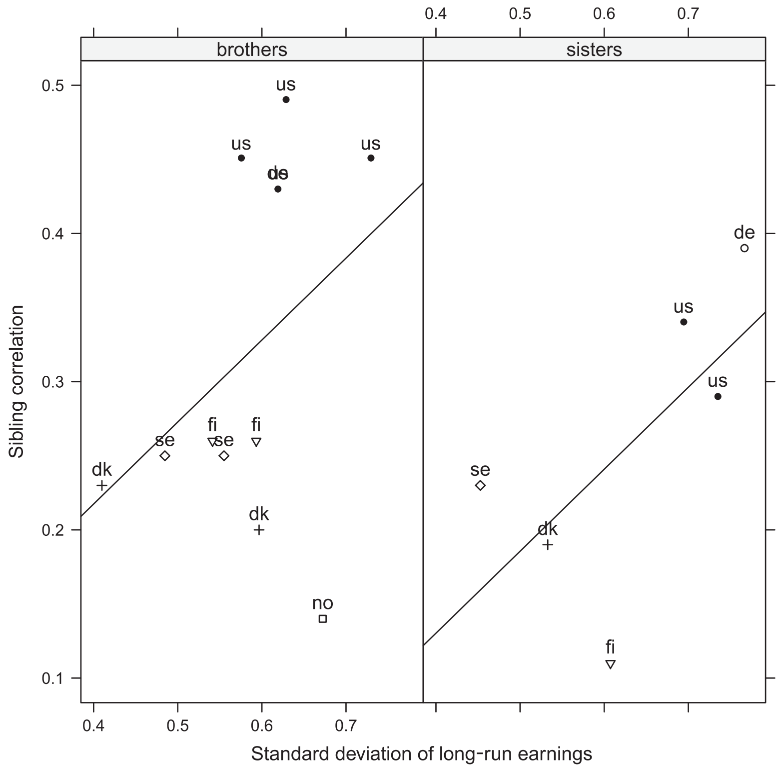


Figure 10.17 Sibling correlation and long-run earnings inequality. *Note: We have plotted on the horizontal axis the sum of the family and individual components, which captures the variance of long-run earnings or income. The vertical axis shows the level of the estimated sibling correlation. Also shown in each panel is the least-squares regression line. Source: See Table 10.9.*

do Germany and the United States but, in the former group of countries, brother correlations are between 0.14 and 0.25. In the latter group, they are 0.43 and as large as 0.49. The regression line for men has all negative deviations for low brother correlations and all positive ones for high correlations, suggesting the least-squares line gives a poor fit. Indeed, if we look at the two "clusters" in each panel—the Nordic countries as one and Germany and the United States as the other—one conclusion may be simply that the Nordic countries differ form the United States and Germany. Thus, although the least-squares line in both panels has a positive slope, it may be premature to talk of a "Great Gatsby" curve for sibling correlations.

There is some evidence about changes across time in brother correlations both in the United States and Sweden. Levine and Mazumder (2007) examined brother correlations in earnings, family income, and hourly wages for two sets of cohorts: those born in 1942–1952 and those born in 1957–1965. The brother correlations in earnings increased from 0.263 to 0.452, in family income from 0.207 to 0.415, and those in hourly wages from 0.277 to 0.472. In no case is the change statistically significant at the conventional 5% level but, taken together, the estimates suggest the importance of family background may have increased quite substantially. By contrast, Björklund et al. (2009) studied change in brother correlations in Sweden starting with cohorts born 1932–1938 and ending in 1962–1968 and found a decline in the importance of family background in the long-run income of men of roughly 13% points. Although the authors are unable to pinpoint the reason for the decline, it coincides with the development of various welfare-state institutions.

We close by noting that, as with intergenerational associations, research on sibling associations should in the future provide more estimates for us to be able to draw robust conclusions about the importance of family background. Apart from the obvious question of why it is that siblings are so similar (what is it that families do?), we would like to see sibling correlations estimated (using the same methods and definitions) for a much wider group of countries than those seven for which we now have information.

We would also like to see rank correlations, not only Pearson correlations, to allow for a full standardization of the marginal distribution when comparing across time, countries, as well as estimates for both women as well as men. A minor point in that regard concerns estimation. Most of the estimates of sibling correlations in Table 10.9 rely on either unbalanced ANOVA or REML to estimate the variance components. Although REML estimates could in principle be defined for data that follow an arbitrary distribution, in practice the likelihood is that of a normal distribution, as a and b are both modeled as conditionally normally distributed variables. Although this may produce reasonably accurate estimates for the log of earnings or income, it is unlikely that REML would produce good estimates if applied to ranks, which are uniformly distributed by definition. Thus, the most feasible way of estimating sibling rank correlations would be to work with pairs of siblings rather than multilevel models.[80]

10.5.5 Other Approaches to Intergenerational Mobility

In this section, we discuss three other approaches to intergenerational mobility. Two, based on occupation and on analysis of surnames, have recently been used to study very long-run trends in intergenerational mobility for which income information is not available. The third concerns an emerging literature on intergenerational links across more than two generations.

Economists have much to learn from sociologists when it comes to the study of intergenerational mobility. The study of the transmission of socioeconomic advantage from generation to generation is one of the core issues in sociology. Empirical research has taken place for almost a hundred years, and the theoretical discussion is also rich. Not

[80] Note that the GMM-based estimates for Sweden reported in Table 10.9 also rely on pairs of brothers.

surprisingly, the available data, the statistical techniques, as well as the possibility to handle large data sets with statistical techniques have improved markedly in the last couple of decades. Hence, the prospects for comparative research based on reasonably comparable data have improved. Nonetheless, comparability is a major concern in the literature that we have come across.

One can distinguish between two strands of intergenerational research in modern sociology.[81] One of them focuses on the relationship between status or prestige attainment of two generations, in general fathers and sons. Occupation is used as the basis to define status and alternative scales that attach status levels to occupations have been suggested in this literature. For example, the famous Duncan status index (Duncan, 1961) used the average education and income of each occupational category. Treiman (1977) has constructed prestige scales from survey data on the average prestige that people attach to various occupations.

The other strand of research defines socioeconomic status in terms of social class but emphasizes that social classes are intrinsically discrete and unordered. Hence, the analytical task is to measure mobility between these classes. The pros and cons of these two approaches to intergenerational mobility have been subject to a more than lively discussion within the sociological research community. Both approaches are prevalent, and each has strong support.[82] The sociological literature on social mobility is far too vast to be reviewed here. One milestone is the monumental book by Erikson and Goldthorpe (1992b), discussed, e.g., in Erikson and Goldthorpe (1992a), Hout and Hauser (1992), and Sorensen (1992). This is a highly mature field that has generated enormous insight into intergenerational mobility.

Indeed, to study long-run changes in intergenerational mobility, class mobility may be the only option. Using census data with names and occupational information, Ferrie (2005) and Long and Ferrie (2007, 2013a) identified father–son pairs by tracking the son of a given father in a later census in the United States and the UK. Ferrie (2005) studied long-run trends in occupational mobility in the United States, and Long and Ferrie (2007, 2013a) compared long-run trends in the United States and the UK. They found that the United States was more fluid in the late nineteenth century than either the UK or the United States in the third quarter of the twentieth century, a finding for which changes in agricultural occupations is central. Their paper generated two critical comments by prominent sociologists, Xie and Killewald (2013) and Hout and Guest (2013), to which they replied (Long and Ferrie, 2013b). Taken together, these papers provide a useful introduction to the use of historical census data to study intergenerational mobility across long periods of time.

[81] Ganzeboom et al. (1991) provide an informative survey of this literature.

[82] For discussions, see, e.g., Ganzeboom et al. (1992, pp. 3–7), Erikson and Goldthorpe (1992a), Hout and Hauser (1992), and Sorensen (1992).

Another emerging strand of literature relies on the fact that surnames convey information on social status. Gregory Clark and collaborators have researched social mobility using data about surnames in Sweden, the United States, England, Japan, India, and China.[83] Güell et al. (2007) and Collado et al. (2012) examined intergenerational mobility in Spain using surnames. This approach has great promise, but it would be more convincing if it could be validated using data that contain either occupation or income so mobility using names could be compared with other, more traditional methods.[84]

Finally, there are a handful of papers that examine intergenerational persistence across more than just two generations. The multigenerational view is lucidly discussed by Mare (2011). Income persistence across multiple generations are estimated at least by Marchon (2008) and Lindahl et al. (2012). In both of those papers, both the parents' and grandparents' income affects offspring income, suggesting that the simple "AR(1)" model of intergenerational transmission is incomplete. These papers provide a perspective that most often goes unremarked on in the intergenerational mobility literature, namely that it relies on a "dynastic" view of parent–child associations. Once grandparents are included in the analysis, care must be taken to distinguish between maternal and paternal grandparents.

10.5.6 Summary and Conclusions

The large literature on intergenerational income mobility that has been surveyed in this section suggests that incomes are, indeed, persistent across generations. What has been learned?

The main lesson is that differences in data (the three "W"s discussed in Section 10.5.1) may account for many of the differences in estimates. Put another way, because of the impact of the combination of life cycle effects and transitory variation in both parent and offspring generations, combined with other data issues, we know surprisingly little either about how income persistence varies across countries, or how it changes within countries over time. We also know very little about exchange mobility (fully standardizing for differences in marginal distributions).

Thus, despite the public prominence of the "Great Gatsby" curve, very little is known about how intergenerational income persistence and mobility vary across countries and how this relates to cross-sectional inequality. More research, using comparable data for multiple countries across multiple cohorts of parents and offspring, is required. With a set of stylized facts about mobility differences and trends, we can then set out to try to explain them.

[83] See, e.g., Clark (2012), Clark (2010), Clark and Cummins (2012), Clark and Ishii (2012), Clark and Landes (2012), Clark et al. (2012), and Hao and Clark (2012).

[84] The UK and U.S. data used by Long and Ferrie (2013a) would be ideal for validating the use of names because the father–son link was initially established using names.

10.6. CONCLUSIONS

This chapter shows that substantial progress has been made in the analysis of income mobility over the last few decades, much of which has been stimulated by the increasing availability of suitable longitudinal data. For within-generation analysis, new household panel and administrative record data abound by comparison with the situation described by Atkinson et al. (1992). For between-generation analysis, the number of suitable data sets has also increased substantially, though not to the same extent (for obvious reasons), and issues of data quality remain relatively more important. Put another way, there has been a more general increase in the availability of good-quality intragenerational income data sets across a relatively large number of rich countries. Good-quality data for analysis of intergenerational income mobility are concentrated among a smaller number of countries. Most longitudinal data (in either context) refers to rich industrialized nations, and it would be interesting to examine the extent to which the patterns found also extend to middle- and low-income countries.

Although the availability of good data has increased substantially, many substantive issues of interest are not yet resolved. Our discussion of within-generation mobility revealed few clear-cut conclusions about whether mobility has been increasing over time or decreasing in particular and whether mobility is greater in one country rather than another. The same can be said in regard to the evidence about income mobility between generations. In short, there remains much scope for systematic empirical analysis.

We have also shown that there has been a substantial increase in the number of mobility measures per se, but the literature has not yet matured in the same way as the measurement of (cross-sectional) income inequality has. Relatively underdeveloped are measures of individual income growth and, especially, of income risk. We would like to see empirical researchers making greater use of the descriptive methods that we have outlined—in order to show the data "as they are" as far as possible—while also carefully selecting summary measures that reflect the mobility concept that is of particular interest. In the intergenerational mobility context, for instance, we have recommended greater use of measures of positional change and less reliance on Beta. More generally, transition matrices are underused.

Our discussion of income mobility has focused on mobility between two time points (with the exception of the discussion of mobility as longer-term inequality reduction). This simplifies the measurement task substantially, but does not remove the need for development of methods for describing individual income *trajectories* over multiple periods. In the intergenerational context, the interest is in not simply the similarities or differences between parents' and children's income, but also the prevalence of "rags to riches and back in three generations" trajectories (for example) relative to other patterns. In the intragenerational context, we are interested not simply in each person's total lifetime income, but also in the patterns of variation over calendar time and age, and how these patterns differ across individuals.

With multidimensional (multiperiod) data, the natural reaction of most analysts is to fit models, with a small number of parameters summarizing the key differences between trajectory patterns. In the Introduction, we briefly cited literatures about the modeling of incomes within or between generations. One of the greatest challenges facing income mobility researchers is to develop tractable models of household income dynamics (not simply earnings dynamics for individuals) both within and between generations. Compared to the field of mobility measurement that we have reviewed in this chapter, mobility modeling is underdeveloped and deserves greater attention in the future.

ACKNOWLEDGMENTS

We wish to thank Daniele Checchi, Philippe Van Kerm, and especially Tony Atkinson and François Bourguignon, for their comments and suggestions on early drafts of this chapter. S. P. Jenkins acknowledges partial financial support for his research from the core funding of the Research Centre on Micro-Social Change by the UK Economic and Social Research Council Core (Grant RES-518-28-001) and the University of Essex.

REFERENCES

Aaberge, R., Björklund, A., Jäntti, M., Palme, M., Pedersen, P., Smith, N., Wennemo, T., 2002. Income inequality and income mobility in the Scandinavian countries compared to the United States. Rev. Income Wealth 48 (4), 443–469.

Aaronson, D., Mazumder, B., 2008. Intergenerational economic mobility in the United States, 1940 to 2000. J. Hum. Resour. 43 (1), 139–172.

Abowd, J.M., Stinson, M.H., 2013. Estimating measurement error in annual job earnings: a comparison of survey and administrative data. Rev. Econ. Stat. 95 (5), 1451–1467.

Aldridge, S., 2001. Social Mobility: A Discussion Paper. Cabinet Office, Performance and Evaluation Unit, London.

Allanson, P.F., 2012. On the characterization and economic evaluation of income mobility as a process of distributional change. J. Econ. Inequal. 10 (4), 505–528.

Altonji, J.G., Dunn, T.A., 1991. Relationships among the family incomes and labor market outcomes of relatives. Res. Labor Econ. 12, 269–310.

Angrist, J.D., Krueger, A.B., 1992. The effect of age at school entry on educational attainment: an application of instrumental variables with moments from two samples. J. Am. Stat. Assoc. 87 (418), 328–336.

Arellano, M., Meghir, C., 1992. Female labour supply and on-the-job search: an empirical model estimated using complementary data sets. Rev. Econ. Stud. 59 (3), 537–559.

Aretz, B., 2013. Gender Differences in German Wage Mobility. Discussion Paper 13–003, Zentrum für Europäische Wirtschaftsforschung, Mannheim. http://hdl.handle.net/10419/68221.

Atkinson, A.B., 1970. On the measurement of inequality. J. Econ. Theory 2 (3), 244–263.

Atkinson, A.B., 1981a. The measurement of economic mobility. In: Eigjelshoven, P.J., van Gemerden, L.J. (Eds.), Essays in Honor of Jan Pen. Het Spectrum, Utrecht.

Atkinson, A.B., 1981b. On intergenerational income mobility in Britain. J. Post Keynes. Econ. 13 (2), 194–218.

Atkinson, A.B., 1983. The measurement of economic mobility. In: Atkinson, A.B. (Ed.), Social Justice and Public Policy. MIT Press, Cambridge, MA, pp. 61–76 (Chapter 3).

Atkinson, A.B., Bourguignon, F., 1982. The comparison of multi-dimensioned distributions of economic status. Rev. Econ. Stud. 49 (2), 183–201.

Atkinson, A.B., Bourguignon, F. (Eds.), 2000. Handbook of Income Distribution, Volume 1. Elsevier, Amsterdam.
Atkinson, A.B., Maynard, A.K., Trinder, C.G., 1983. Parents and Children: Incomes in Two Generations. Heinemann Educational Books, London.
Atkinson, A.B., Bourguignon, F., Morrisson, C., 1992. Empirical Studies of Earnings Mobility. Fundamentals in Pure and Applied Economics. Harwood Academic Publisher, Chur.
Auten, G., Gee, G., 2009. Income mobility in the United States: new evidence from tax data. Natl. Tax J. 62 (2), 301–328.
Auten, G., Gee, G., Turner, N., 2013. Income inequality, mobility, and turnover at the top in the US, 1987–2010. Am. Econ. Rev. 103 (3), 168–172.
Ayala, L., Sastre, M., 2008. The structure of income mobility: empirical evidence from five UE countries. Empir. Econ. 35 (4), 451–473.
Bach, S., Corneo, G., Steiner, V., 2009. From bottom to top: the entire income distribution in Germany. Rev. Income Wealth 55 (2), 303–330.
Baker, M., Solon, G., 2003. Earnings dynamics and inequality among Canadian men, 1976–1992: evidence from longitudinal income tax records. J. Labor Econ. 21 (2), 267–288.
Bane, M., Ellwood, D., 1986. Slipping into and out of poverty: the dynamics of spells. J. Hum. Resour. 21 (1), 1–23.
Bartels, C., Bönke, T., 2013. Can households and welfare states mitigate rising earnings instability? Rev. Income Wealth 59 (2), 250–282.
Bartholomew, D.J., 1973. Stochastic Models for Social Processes, second ed. Wiley, Chichester.
Bayaz-Ozturk, G., Burkhauser, R.V., Couch, K.A., 2014. Consolidating the evidence on income mobility in the Western states of Germany and the U.S. from 1984–2006. Econ. Inq. 52 (1), 431–443.
Becker, G.S., 1989. On the economics of the family: reply to a skeptic. Am. Econ. Rev. 79 (3), 514–518.
Becker, G.S., Tomes, N., 1979. An equilibrium theory of the distribution of income and intergenerational mobility. J. Polit. Econ. 87 (6), 1153–1189.
Becker, G.S., Tomes, N., 1986. Human capital and the rise and fall of families. J. Labor Econ. 4 (3), S1–S39.
Benabou, R., Ok, E.A., 2000. Mobility as Progressivity: Ranking Income Processes According to Equality of Opportunity. Working Paper w8431, National Bureau of Economic Research, Cambridge, MA. http://www.nber.org/papers/w8431.
Benabou, R., Ok, E.A., 2001. Social mobility and the demand for redistribution: the POUM hypothesis. Q. J. Econ. 116 (2), 447–487.
Bhattacharya, D., Mazumder, B., 2011. A nonparametric analysis of black-white differences in intergenerational income mobility in the United States. Quant. Econ. 2 (3), 335–379.
Biewen, M., 2002. Bootstrap inference for inequality, mobility and poverty measurement. J. Econ. 108 (2), 317–342.
Björklund, A., Chadwick, L., 2003. Intergenerational income mobility in permanent and separated families. Econ. Lett. 80 (2), 239–246.
Björklund, A., Jäntti, M., 1997. Intergenerational income mobility in Sweden compared to the United States. Am. Econ. Rev. 87 (4), 1009–1018.
Björklund, A., Jäntti, M., 2000. Intergenerational mobility of socio-economic status in comparative perspective. Nord. J. Polit. Econ. 26 (1), 3–32.
Björklund, A., Jäntti, M., 2009. Intergenerational income mobility and the role of family background. In: Salverda, W., Nolan, B., Smeeding, T.M. (Eds.), Oxford Handbook of Economic Inequality. Oxford University Press, Oxford, pp. 491–521 (Chapter 20).
Björklund, A., Eriksson, T., Jäntti, M., Raaum, O., Österbacka, E., 2002. Brother correlations in earnings in Denmark, Finland, Norway and Sweden compared to the United States. J. Popul. Econ. 15 (4), 757–772.
Björklund, A., Eriksson, T., Jäntti, M., Raaum, O., Österbacka, E., 2004. Family structure and labour market success: the influence of siblings and birth order on the earnings of young adults in Norway, Finland and Sweden. In: Corak, M. (Ed.), Generational Income Mobility in North America and Europe. Cambridge University Press, Cambridge, pp. 207–225 (Chapter 9).

Björklund, A., Jäntti, M., Lindquist, M.J., 2007a. Family Background and Income During the Rise of the Welfare State: Brother Correlations in Income for Swedish Men Born 1932–1968. IZA Discussion Paper 3000, IZA, Bonn. http://ftp.iza.org/dp3000.pdf.

Björklund, A., Jäntti, M., Solon, G., 2007b. Nature and nurture in the intergenerational transmission of socioeconomic status: evidence from Swedish children and their biological and rearing parents. B.E. J. Econ. Anal. Policy 7 (2), Article 4.

Björklund, A., Jäntti, M., Lindquist, M.J., 2009. Family background and income during the rise of the welfare state: brother correlations in income for Swedish men born 1932–1968. J. Public Econ. 93 (5–6), 671–680.

Björklund, A., Lindahl, L., Lindquist, M.J., 2010. What more than parental income, education and occupation? An exploration of what Swedish siblings get from their parents. B.E. J. Econ. Anal. Policy 10 (1), Article 102.

Björklund, A., Roine, J., Waldenström, D., 2012. Intergenerational top income mobility in Sweden: capitalist dynasties in the land of equal opportunity? J. Public Econ. 96 (5), 474–484.

Black, S.E., Devereux, P.J., 2011. Recent developments in intergenerational mobility. In: Ashenfelter, O., Card, D. (Eds.), In: Handbook of Labor Economics, vol. 4B. Elsevier, Amsterdam, pp. 1487–1541 (Chapter 16).

Blanden, J., 2013. Cross-country rankings in intergenerational mobility: a comparison of approaches from economics and sociology. J. Econ. Surv. 27 (1), 38–73.

Blanden, J., Machin, S., 2007. Recent Changes in Intergenerational Mobility in Britain. The Sutton Trust, London.

Blanden, J., Machin, S., 2008. Up and down the generational income ladder in Britain: past changes and future prospects. Natl. Inst. Econ. Rev. 205 (1), 101–106.

Blanden, J., Gregg, P., Macmillan, L., 2010. Intergenerational Persistence in Income and Social Class: The Impact of Within-Group Inequality. Working Paper 10/230, Centre for Market and Public Organisation, University of Bristol. http://www.bristol.ac.uk/cmpo/publications/papers/2010/wp230.pdf.

Blanden, J., Gregg, P., Macmillan, L., 2013. Intergenerational persistence in income and social class: the effect of within-group inequality. J. R. Stat. Soc. Ser. A Stat. Soc. 176 (2), 541–563.

Blundell, R., Pistaferri, L., Preston, I., 2008. Consumption inequality and partial insurance. Am. Econ. Rev. 98 (5), 1887–1921.

Böheim, R., Jenkins, S.P., 2006. A comparison of current and annual measures of income in the British Household Panel Survey. J. Off. Stat. 22 (4), 733–758.

Böhlmark, A., Lindquist, M.J., 2006. Life-cycle variations in the association between current and lifetime income: replication and extension for Sweden. J. Labor Econ. 24 (4), 879–896.

Bonke, J., Hussain, M.A., Munk, M.D., 2005. A Comparison of Danish and International Findings on Intergenerational Earnings Mobility. Danish National Institute of Social Research, Copenhagen. unpublished manuscript, https://www.pisa2012.dk/Files/Filer/SFI/Pdf/Arbejdspapirer/Arbejdspapirer/2005_11_WorkingPaper.pdf.

Bossert, W., Chakravarty, S., D'Ambrosio, C., 2012. Poverty and time. J. Econ. Inequal. 10 (2), 145–162.

Bound, J., Brown, C., Mathiowetz, N., 2001. Measurement error in survey data. In: Heckman, J.J., Leamer, E. (Eds.), Handbook of Econometrics, vol. 5. Elsevier Science, Amsterdam, pp. 3707–3745.

Bourguignon, F., 2011. Non-anonymous growth incidence curves, income mobility and social welfare dominance. J. Econ. Inequal. 9 (4), 605–627.

Bowles, S., Gintis, H., 2002. The inheritance of inequality. J. Econ. Perspect. 16 (3), 3–30.

Bradbury, K., 2011. Trends in U.S. Family Income Mobility, 1969–2006. Working Paper 11-10, Federal Reserve Bank of Boston, Boston, MA. http://www.bos.frb.org/economic/wp/wp2011/wp1110.pdf.

Bradbury, K., Katz, J., 2002. Are lifetime incomes growing more unequal? Looking at new evidence on family income mobility. Fed. Reserve Bank Boston Reg. Rev. Q4, 3–5.

Bratberg, E., Nielsen, Ø.A., Vaage, K., 2007. Trends in intergenerational mobility across offspring's earnings distribution in Norway. Ind. Relat. 46 (1), 112–128.

Bratsberg, B., Røed, K., Raaum, O., Naylor, R., Jäntti, M., Eriksson, T., Österbacka, E., 2007. Nonlinearities in intergenerational earnings mobility: consequences for cross-country comparisons. Econ. J. 117 (519), C72–C92.

Breen, R., Moisio, P., 2004. Poverty dynamics corrected for measurement error. J. Econ. Inequal. 2 (3), 171–191.
Buchinsky, M., Hunt, J., 1999. Wage mobility in the United States. Rev. Econ. Stat. 81 (3), 351–368.
Burkhauser, R., Couch, K., 2009. Intragenerational inequality and intertemporal mobility. In: Salverda, W., Nolan, B., Smeeding, T.M. (Eds.), Oxford Handbook of Economic Inequality. Oxford University Press, Oxford, pp. 522–548 (Chapter 21).
Burkhauser, R.V., Poupore, J., 1997. A cross-national comparison of permanent income inequality. Rev. Econ. Stat. 79 (1), 10–17.
Burkhauser, R.V., Holtz Eakin, D., Rhody, S.E., 1997. Labor earnings, mobility, and equality in the United States and Germany during the growth years of the 1980s. Int. Econ. Rev. 38 (4), 775–794.
Burkhauser, R.V., Holtz Eakin, D., Rhody, S.E., 1998. Mobility and inequality in the 1980s: a cross-national comparison of the United States and Germany. In: Jenkins, S.P., Kapteyn, A., van Praag, B. M.S. (Eds.), The Distribution of Welfare and Household Production: International Perspectives. Cambridge University Press, Cambridge, pp. 111–175 (Chapter 6).
Burkhauser, R.V., Feng, S., Jenkins, S.P., Larrimore, J., 2011. Estimating trends in us income inequality using the Current Population Survey: the importance of controlling for censoring. J. Econ. Inequal. 9 (1), 393–415.
Buron, L., 1994. A Study of the Magnitude and Determinants of Intergenerational Earnings Mobility. Ph.D. thesis, University of Wisconsin-Madison, Madison, WI.
Cantó, O., Gradín, C., Del Río, C., 2006. Poverty statics and dynamics: does the accounting period matter? Int. J. Soc. Welf. 15, 209–218.
Cappellari, L., Jenkins, S.P., 2014. Earnings and Labor Market Volatility in Britain. Labour Econ. forthcoming. http://dx.doi.org/10.1016/j.labeco.2014.03.012.
Celik, S., Juhn, C., McCue, K., Thompson, J., 2012. Recent trends in earnings volatility: evidence from survey and administrative data. B.E. J. Econ. Anal. Policy. 12 (2), Article 1.
Cervini-Plá, M., 2009. Measuring Intergenerational Earnings Mobility in Spain: A Selection-Bias-Free Approach. Working Paper 0904, Department of Applied Economics at Universitat Autonoma of Barcelona. http://EconPapers.repec.org/RePEc:uab:wprdea:wpdea0904.
Chadhuri, S., Ravallion, M., 1994. How well do static indicators identify the chronically poor? J. Public Econ. 53 (3), 367–394.
Chadwick, L., Solon, G., 2002. Intergenerational income mobility among daughters. Am. Econ. Rev. 92 (1), 335–344.
Chakravarty, S.R., 1984. Normative indices for measuring social mobility. Econ. Lett. 15 (1–2), 175–180.
Chakravarty, S., Dutta, J., Weymark, J., 1985. Ethical indices of income mobility. Soc. Choice Welf. 2 (2), 1–21.
Chau, T.W., 2012. Intergenerational income mobility revisited: estimation with an income dynamic model with heterogeneous age profile. Econ. Lett. 117 (3), 770–773.
Checchi, D., Ichino, A., Rustichini, A., 1999. More equal but less mobile? Education financing and intergenerational mobility in Italy and in the US. J. Public Econ. 74 (3), 351–393.
Chen, W.-H., 2009. Cross-national differences in income mobility: evidence from Canada, the United States, Great Britain and Germany. Rev. Income Wealth 55 (1), 75–100.
Chen, T., Couch, K.A., 2013. Permanent and transitory inequality and intragenerational mobility. Econ. Lett. 120 (2), 200–202.
Citro, C.F., Michael, R.T., 1995. Measuring Poverty: A New Approach. National Academy Press, Washington, DC.
Clark, G., 2010. Measuring inequality through the strength of inheritance. Curr. Anthropol. 51 (1), 101–102.
Clark, G., 2012. What is the True Rate of Social Mobility in Sweden? A Surname Analysis, 1700–2012. UC Davis. unpublished manuscript, http://www.econ.ucdavis.edu/faculty/gclark/research.html.
Clark, G., Cummins, N., 2012. What is the True Rate of Social Mobility? Surnames and Social Mobility, England 1800–2012. UC Davis. unpublished manuscript, http://www.econ.ucdavis.edu/faculty/gclark/research.html.

Clark, G., Ishii, T., 2012. Social Mobility in Japan, 1868–2012: The Surprising Persistence of the Samurai. UC Davis. unpublished manuscript, http://www.econ.ucdavis.edu/faculty/gclark/research.html.
Clark, G., Landes, Z., 2012. Caste Versus Class: Social Mobility in India, 1860–2012. UC Davis. unpublished manuscript, http://www.econ.ucdavis.edu/faculty/gclark/research.html.
Clark, G., Marcin, D., Abu-Sneneh, F., Chow, W.M., Jung, K.M., Marek, A.M., Williams, K.M., 2012. Social Mobility Rates in the USA, 1920–2010: A Surname Analysis. UC Davis. unpublished manuscript, http://www.econ.ucdavis.edu/faculty/gclark/research.html.
Collado, M.D., Ortuño-Ortín, I., Romeu, A., 2012. Long-run Intergenerational Social Mobility and the Distribution of Surnames. University of Alicante, Alicante. Unpublished paper, http://www.eco.uc3m.es/personal/iortuno/mob11Febrero.pdf.
Conlisk, J., 1974. Can equalization of opportunity reduce social mobility? Am. Econ. Rev. 64 (1), 80–90.
Conlisk, J., 1977. An exploratory model of the size distribution of income. Econ. Inq. 15 (3), 345–366.
Conlisk, J., 1984. Four invalid propositions about equality, efficiency, and intergenerational transfers through schooling. J. Hum. Resour. 19 (1), 3–21.
Corak, M., 2006. Do Poor Children Become Poor Adults? Lessons for Public Policy from a Cross Country Comparison of Generational Earnings Mobility. Discussion Paper 1993, IZA, Bonn.
Corak, M., 2013a. Income inequality, equality of opportunity, and intergenerational mobility. J. Econ. Perspect. 27 (3), 79–102.
Corak, M., 2013b. Inequality from generation to generation: the United States in comparison. In: Rycroft, R.S. (Ed.), The Economics of Inequality, Poverty, and Discrimination in the 21st Century. ABC-CLIO, Santa Barbara, CA (Chapter 6).
Corak, M., Heisz, A., 1999. The intergenerational earnings and income mobility of Canadian men: evidence from longitudinal income tax data. J. Hum. Resour. 34 (3), 504–556.
Corak, M., Lindquist, M.J., Mazumder, B., 2013. A comparison of upward and downward intergenerational mobility in Canada, Sweden and the United States. In: Paper presented at the 2013 EALE Conference, Turin, http://www.eale.nl/Conference2013/program/Parallel%20session%20A/add215310_konuoeQdIq.pdf.
Couch, K.A., Dunn, T.A., 1997. Intergenerational correlations in labor market status: a comparison of the United States and Germany. J. Hum. Resour. 32 (1), 210–232.
Couch, K.A., Lillard, D.R., 1998. Sample selection rules and the intergenerational correlation of earnings. Labour Econ. 5 (3), 313–329.
Couch, K.A., Lillard, D.R., 2004. Nonlinearities in intergenerational mobility: a comparison of Germany and the United States. In: Corak, M. (Ed.), Generational Income Mobility in North America and Europe. Cambridge University Press, Cambridge, pp. 190–206 (Chapter 8).
Cowell, F.A., 1985. Measures of distributional change: an axiomatic approach. Rev. Econ. Stud. 52 (1), 135–151.
Cowell, F.A., Flachaire, E., 2011. Measuring Mobility. Discussion Paper PEP09, STICERD, London School of Economics. http://sticerd.lse.ac.uk/dps/pep/pep09.pdf.
Cowell, F.A., Schluter, C., 1999. Income mobility: a robust approach. In: Silber, J. (Ed.), Income Inequality Measurement: From Theory to Practice. Kluwer, Dewenter.
Creedy, J., Halvorsen, E., Thoresen, T.O., 2013. Inequality comparisons in a multi-period framework: the role of alternative welfare metrics. Rev. Income Wealth 59 (2), 235–249.
Cunha, F., Heckman, J., Navarro, S., 2005. Separating uncertainty from heterogeneity in life cycle earnings. Oxf. Econ. Pap. 57 (2), 191–261.
D'Agostino, M., Dardanoni, V., 2009a. The measurement of rank mobility. J. Econ. Theory 144 (4), 1783–1803.
D'Agostino, M., Dardanoni, V., 2009b. What's so special about Euclidean distance? A characterization with applications to mobility and spatial voting. Soc. Choice Welf. 33 (2), 211–233.
Dahl, M., DeLeire, T., 2008. The Association Between Children's Earnings and Fathers' Lifetime Earnings: Estimates using Administrative Data. Discussion Paper 1342-08, Institute for Research on Poverty, University of Wisconsin-Madison. http://www.irp.wisc.edu/publications/dps/pdfs/dp134208.pdf.
Dahl, M., DeLeire, T., Schwabisch, J.A., 2011. Estimates of year-to-year volatility in earnings and in household incomes from administrative, survey, and matched data. J. Hum. Resour. 46 (4), 750–774.

Dardanoni, V., 1993. Measuring social mobility. J. Econ. Theory 61 (2), 372–394.
Dearden, L., Machin, S., Reed, H., 1997. Intergenerational mobility in Britain. Econ. J. 107 (1), 47–66.
DeBacker, J., Heim, B., Panousi, V., Ramnath, S., Vidangos, I., 2013. Rising inequality: transitory or permanent? New evidence from a panel of U.S. tax returns. Brookings Papers on Economic Activity 46 (1), 67–122.
Demuynck, T., Van de gaer, D., 2012. Inequality adjusted income growth. Economica 79 (316), 747–765.
Department for Work and Pensions, 2009. Low Income Dynamics, 1991–2007. Department for Work and Pensions, London.
Dickens, R., McKnight, A., 2008. Changes in Earnings Inequality and Mobility in Great Britain 1978/9–2005/6. CASEpaper 132, Centre for the Analysis of Social Exclusion, London School of Economics, London. http://sticerd.lse.ac.uk/dps/case/cp/CASEpaper132.pdf.
Doris, A., O'Neill, D., Sweetman, O., 2013. Identification of the covariance structure of earnings using the GMM estimator. J. Econ. Inequal. 11 (3), 343–372.
Dragoset, L.M., Fields, G.S., 2006. U.S. Earnings Mobility: Comparing Survey-Based and Administrative-Based Estimates. ECINEQ Working Paper 2006-55, Society for the Study of Economic Inequality, Milan. http://www.ecineq.org/milano/WP/ECINEQ2006-55.pdf.
Duclos, J.Y., Araar, A., Giles, J., 2010. Chronic and transient poverty: measurement and estimation, with evidence from China. J. Dev. Econ. 91 (2), 266–277.
Duncan, O.D., 1961. A socio-economic index for all occupations and properties and characteristics of occupational status. In: Reiss, A.J.J. (Ed.), Occupations and Social Status. Free Press of Glencoe, New York.
Duncan, G., Gustafsson, B., Hauser, R., Shmauss, G., Messinger, H., Muffels, R., Nolan, B., Ray, J.-C., 1993. Poverty dynamics in eight countries. J. Popul. Econ. 6 (3), 215–234.
Dunn, C.E., 2007. The intergenerational transmission of lifetime earnings: evidence from Brazil. B.E. J. Econ. Anal. Policy 7 (2), 1–42.
Dutta, I., Roope, L., Zank, H., 2013. On intertemporal poverty measures: the role of affluence and want. Soc. Choice Welf. 41 (4), 721–740.
Dynan, K.E., Elmendorf, D., Sichel, D.E., 2012. The evolution of household income volatility. B.E. J. Econ. Anal. Policy 12 (2), Article 3.
Eberharter, V.V., 2013. The intergenerational dynamics of social inequality—empirical evidence from Europe and the United States. In: Paper presented at the ECINEQ Conference, Bari, July 2013. http://www.ecineq.org/ecineq_bari13/FILESxBari13/CR2/p118.pdf.
Ehling, M., Rendtel, U. (Eds.), 2004. Harmonisation of panel surveys and data quality. The Change from Input Harmonization to Ex-post Harmonization in National Samples of the European Community Household Panel—Implications on Data Quality, Statistisches Bundesamt, Wiesbaden, CHINTEX.
Eide, E., Showalter, M., 1999. Factors affecting the transmission of earnings across generations: a quantile regression approach. J. Hum. Resour. 34 (2), 253–267.
Erikson, R., Goldthorpe, J.H., 1992a. The CASMIN project and the American Dream. Eur. Sociol. Rev. 8 (3), 283–305.
Erikson, R., Goldthorpe, J.H., 1992b. The Constant Flux—A Study of Class Mobility in Industrial Societies. Clarendon Press, Oxford.
Erikson, R., Goldthorpe, J.H., 2010. Income and class mobility between generations in Great Britain: the problem of divergent findings from the data-sets of birth cohort studies. Br. J. Sociol. 61 (2), 211–230.
Eriksson, T., Zhang, Y., 2012. The role of family background for earnings in rural china. Front. Econ. China 7 (3), 465–477.
Expert Group on Household Income Statistics (The Canberra Group), 2001. Final Report and Recommendations. Ottawa, http://www.lisproject.org/links/canberra/finalreport.pdf.
Ferrie, J.P., 2005. History lessons: the end of American exceptionalism? Mobility in the United States since 1850. J. Econ. Perspect. 19 (3), 199–215.
Fertig, A.M., 2003. Trends in intergenerational earnings mobility in the U.S. J. Income Distribution 12 (3–4), 108–130.
Fields, G.S., 2006. The many facets of economic mobility. In: McGillivray, M. (Ed.), Inequality, Poverty, and Well-Being. Palgrave Macmillan, Houndmills, pp. 123–142.

Fields, G.S., 2010. Does income mobility equalize longer-term incomes? New measures of an old concept. J. Econ. Inequal. 8 (4), 409–427.
Fields, G.S., Ok, E.A., 1996. The meaning and measurement of income mobility. J. Econ. Theory 71 (2), 349–377.
Fields, G.S., Ok, E.A., 1999a. The measurement of income mobility: an introduction to the literature. In: Silber, J. (Ed.), Handbook of Income Inequality Measurement. Series on Recent Economic Thought, vol. 71. Kluwer Academic Publishers, Boston, pp. 557–598 (Chapter 19).
Fields, G.S., Ok, E.A., 1999b. Measuring movement of incomes. Economica 66 (264), 455–471.
Fields, G.S., Leary, J.B., Ok, E.A., 2002. Stochastic dominance in mobility analysis. Econ. Lett. 75 (3), 333–339.
Fields, G.S., Cichello, P.L., Freije, S., Menédez, M., Newhouse, D., 2003. For richer or for poorer? Evidence from Indonesia, South Africa, Spain, and Venezuela. J. Econ. Inequal. 1 (1), 67–99.
Fitzgerald, J., Gottschalk, P., Moffitt, R., 1998. An analysis of attrition in panel data: the Michigan Panel Study of Income Dynamics. J. Hum. Resour. 33 (2), 251–299.
Formby, J.P., Smith, W.J., Zheng, B., 2004. Mobility measurement, transition matrices and statistical inference. J. Econ. 120 (1), 181–205.
Foster, J.E., 2009. A class of chronic poverty measures. In: Addison, T., Hulme, D., Kanbur, R. (Eds.), Poverty Dynamics: Interdisciplinary Perspectives. Oxford University Press, Oxford, pp. 59–76.
Foster, J.E., Greer, J., Thorbecke, E., 1984. A class of decomposable poverty indices. Econometrica 52 (3), 761–766.
Frick, J.R., Jenkins, S.P., Lillard, D.R., Lipps, O., Wooden, M., 2007. The Cross-National Equivalent File (CNEF) and its member country household panel studies. Schmollers Jahrbuch: Journal of Applied Social Sciences Studies 127 (4), 627–654.
Friedman, M., 1962. Capitalism and Freedom. Chicago University Press, Chicago.
Fritzell, J., 1990. The dynamics of income distribution: economic mobility in Sweden in comparison with the United States. Soc. Sci. Res. 19 (1), 17–46.
Galton, F., 1886. Regression towards mediocrity in hereditary stature. J. Anthropol. Inst. 15, 246–263.
Gangl, M., 2005. Income inequality, permanent incomes, and income dynamics. Comparing Europe to the United States. Work. Occup. 37 (2), 140–162.
Ganzeboom, H.B., Treiman, D.J., Ultee, W.C., 1991. Comparative intergenerational stratification research: three generations and beyond. Annu. Rev. Sociol. 17, 277–302.
Ganzeboom, H., Graaf, P.M.D., Treiman, D.J., Leeuw, J.D., 1992. A standard international socio-economic index of occupational status. Soc. Sci. Res. 21 (1), 1–56.
Gibbons, M., 2010. Income and Occupational Intergenerational Mobility in New Zealand. Working Paper 10/06New Zealand Treasury, Wellington. http://www.treasury.govt.nz/publications/research-policy/wp/2010/10-06.
Gittleman, M., Joyce, M., 1999. Have family income mobility patterns changed? Demography 36 (3), 299–314.
Goldberger, A.S., 1989. Economic and mechanical models of intergenerational transmission. Am. Econ. Rev. 79 (3), 504–513.
Goldthorpe, J.H., 2013. Entry of the economists, the confusion of politicians and the limits of educational policy. J. Soc. Policy 42 (3), 431–450.
Gottschalk, P., Huynh, M., 2010. Are earnings inequality and mobility overstated? The impact of non-classical measurement error. Rev. Econ. Stat. 92, 302–315.
Gottschalk, P., Moffitt, R., 1994. The growth of earnings instability in the U.S. labor market. Brook. Pap. Econ. Act. 1994 (2), 217–272.
Gottschalk, P., Moffitt, R.A., 2009. The rising instability of U.S. earnings. J. Econ. Perspect. 23, 3–24.
Gottschalk, P., Spolaore, E., 2002. On the evaluation of economic mobility. Rev. Econ. Stud. 69, 191–208.
Gouskova, E., Chiteji, N., Stafford, F., 2010. Estimating the intergenerational persistence of lifetime earnings with life course matching: evidence from the PSID. Labour Econ. 17 (3), 592–597.
Gradín, C., Del Río, C., Cantó, O., 2012. Measuring poverty accounting for time. Rev. Income Wealth 58 (3), 330–354.

Grawe, N.D., 2004a. Intergenerational mobility for whom? The experience of high- and low-earnings sons in international perspective. In: Corak, M. (Ed.), Generational Income Mobility in North America and Europe. Cambridge University Press, Cambridge, pp. 58–89 (Chapter 4).
Grawe, N.D., 2004b. Reconsidering the use of nonlinearities in intergenerational earnings mobility as a test for credit constraints. J. Hum. Resour. 39 (3), 813–827.
Grawe, N.D., 2006. Lifecycle bias in estimates of intergenerational earnings persistence. Labour Econ. 13 (5), 519–664.
Gregg, P., Vittori, C., 2009. Earnings mobility in Europe: Global and disaggregate measures. CMPO, University of Bristol, unpublished paper.
Grimm, M., 2007. Removing the anonymity axiom in assessing pro-poor growth. J. Econ. Inequal. 5 (2), 179–197.
Güell, M., Mora, J.V.R., Telmer, C., 2007. Intergenerational Mobility and the Informative Content of Surnames. CEP Discussion Paper 0810, Centre for Economic Performance, LSE. http://ideas.repec.org/p/cep/cepdps/dp0810.html.
Guvenen, F., 2009. An empirical investigation of labor income processes. Rev. Econ. Dyn. 12 (1), 58–79.
Hacker, J.S., Jacobs, E., 2008. The Rising Instability of American Family Incomes, 1969–2004. Evidence from the Panel Study of Income Dynamics. Briefing Paper 213, Economic Policy Institute, Washington, DC. http://www.epi.org/content.cfm/bp213.
Haider, S.J., 2001. Earnings instability and earnings inequality of males in the United States: 1967–1991. J. Labor Econ. 19 (4), 799–836.
Haider, S., Solon, G., 2003. Life-Cycle Variation in the Association Between Current and Lifetime Earnings. Unpublished manuscript, Department of Economics, Michigan State University, East Lansing, MI.
Haider, S., Solon, G., 2006. Life-cycle variation in the association between current and lifetime earnings. Am. Econ. Rev. 96 (4), 1308–1320.
Hansen, M.N., 2010. Change in intergenerational economic mobility in Norway. Conventional versus joint classifications of economic origin. J. Econ. Inequal. 8 (2), 133–151.
Hao, Y., Clark, G., 2012. Social Mobility in China, 1645–2012: A Surname Study. UC Davis. unpublished manuscript, http://www.econ.ucdavis.edu/faculty/gclark/research.html.
Hart, P.E., 1976. The comparative statics and dynamics of income distributions. J. R. Stat. Soc. Ser. A 139 (1), 108–125.
Hassler, J., Rodrigues, J.V., Zeira, J., 2007. Inequality and mobility. J. Econ. Growth 12 (3), 221–259.
Hertz, T., 2005. Rags, riches, and race: the intergenerational economic mobility of black and white families in the United States. In: Bowles, S., Gintis, H., Osborne, M. (Eds.), Unequal Chances: Family Background and Economic Success. Russell Sage Foundation, New York, pp. 165–191 (Chapter 5).
Hertz, T., 2007. Trends in the intergenerational elasticity of family income in the United States. Ind. Relat. 46 (1), 22–50.
Holmlund, H., Lindahl, M., Plug, E., 2011. The causal effect of parents' schooling on children's schooling: a comparison of estimation methods. J. Econ. Lit. 49 (3), 615–651.
Hout, M., Guest, A.M., 2013. Intergenerational occupational mobility in Great Britain and the United States since 1850: comment. Am. Econ. Rev. 103 (5), 2021–2040.
Hout, M., Hauser, R.M., 1992. Symmetry and hierarchy in social mobility: a methodological analysis of the CASMIN model of class mobility. Eur. Sociol. Rev. 8 (3), 239–266.
Hungerford, T.L., 1993. Income mobility in the Seventies and Eighties. Rev. Income Wealth 39, 403–417.
Hungerford, T.L., 2011. How income mobility affects income inequality: US evidence in the 1980s and 1990s. J. Income Distribution 20 (1), 83–103.
Inoue, A., Solon, G., 2010. Two-sample instrumental variables estimators. Rev. Econ. Stat. 92 (3), 557–561.
Jalan, J., Ravallion, M., 1998. Transient poverty in postreform rural China. J. Comp. Econ. 26 (2), 338–357.
Jäntti, M., Lindahl, L., 2012. On the variability of income within and across generations. Econ. Lett. 117 (1), 165–167.
Jäntti, M., Kanbur, R., Nyyssölä, M., Pirttilä, J., 2014. Poverty and welfare measurement on the basis of prospect theory. Rev. Income Wealth 60 (1), 182–205.
Jarvis, S., Jenkins, S.P., 1998. How much income mobility is there in Britain? Econ. J. 108 (447), 428–443.

Jenderny, K., 2013. Mobility of Top Incomes in Germany. Discussion Paper 2013/7, School of Business and Economics, Free University of Berlin. http://www.wiwiss.fu-berlin.de/institute/finanzen/schoeb/lehrstuhl/jenderny/discpaper7_2013.pdf?1370475263.

Jenkins, S.P., 1987. Snapshots versus movies: 'lifecycle biases' and the estimation of intergenerational earnings inheritance. Eur. Econ. Rev. 31 (5), 1149–1158.

Jenkins, S.P., 1994. 'Social welfare function' measures of horizontal inequity. In: Eichhorn, W. (Ed.), Models and Measurement of Welfare and Inequality. Springer-Verlag, Berlin, pp. 725–751.

Jenkins, S.P., 2000. Modelling household income dynamics. J. Popul. Econ. 13 (4), 529–567.

Jenkins, S.P., 2011a. Changing Fortunes: Income Mobility and Poverty Dynamics in Britain. Oxford University Press, Oxford.

Jenkins, S.P., 2011b. Has the instability of personal incomes been increasing? Natl. Inst. Econ. Rev. 218 (1), R33–R43.

Jenkins, S.P., Van Kerm, P., 2006. Trends in income inequality, pro-poor income growth, and income mobility. Oxf. Econ. Pap. 58 (3), 531–548.

Jenkins, S.P., Van Kerm, P., 2009. The measurement of economic inequality. In: Salverda, W., Nolan, B., Smeeding, T.M. (Eds.), The Oxford Handbook on Economic Inequality. Oxford University Press, Oxford, pp. 40–67.

Jenkins, S.P., Van Kerm, P., 2011. Trends in Individual Income Growth: Measurement Methods and British Evidence. ISER Working Paper Series 2011-06, Institute for Social and Economic Research, University of Essex, Colchester. http://ideas.repec.org/p/ese/iserwp/2011-06.html.

Jenkins, S.P., Van Kerm, P., 2014. The relationship between EU indicators of persistent and current poverty. Soc. Indic. Res. 116 (2), 611–638.

Juhn, C., McCue, K., 2010. Comparing Estimates of Earnings Instability Based on Survey and Administrative Reports. Working Paper CES 2010-15, Bureau of the Census, Washington, DC. ftp://ftp2.census.gov/ces/wp/2010/CES-WP-10-15.pdf.

Kanbur, S.M.R., Stiglitz, J.E., 1986. Intergenerational Mobility and Dynastic Inequality. Research Memorandum 324, Econometric Research Program, Princeton University, Princeton, NJ. http://www.princeton.edu/~erp/ERParchives/archivepdfs/M324.pdf.

King, M.A., 1983. An index of inequality: with applications to horizontal equity and social mobility. Econometrica 51 (1), 99–115.

Koenker, R., 2005. Quantile Regression. Cambridge University Press, Cambridge, UK.

Kopczuk, W., Saez, E., Song, J., 2010. Earnings inequality and mobility in the United States: evidence from Social Security data since 1937. Q. J. Econ. 125 (1), 91–128.

Krueger, A.B., 2012. The rise and consequences of inequality in the United States, http://www.whitehouse.gov/sites/default/files/krueger_cap_speech_final_remarks.pdf.

Lee, C.-I., Solon, G.M., 2009. Trends in intergenerational income mobility. Rev. Econ. Stat. 91 (4), 766–772.

Lee, S., Linton, O., Whang, Y.-J., 2009. Testing for stochastic monotonicity. Econometrica 77 (2), 585–602.

Lefranc, A., 2011. Educational Expansion, Earnings Compression and Changes in Intergenerational Economic Mobility: Evidence from French Cohorts, 1931–1976. THEMA Working Paper 2011-11, Université de Cergy-Pontoise. http://www.u-cergy.fr/thema/repec/2011-11.pdf.

Lefranc, A., Pistolesi, N., Trannoy, A., 2009. Equality of opportunity and luck: definitions and testable conditions, with an application to income in France. J. Public Econ. 93 (11–12), 1189–1207.

Lefranc, A., Ojima, F., Yoshida, T., 2013. Intergenerational earnings mobility in Japan among sons and daughters: levels and trends. J. Popul. Econ. 27 (1), 91–134.

Leigh, A., 2007. Intergenerational mobility in Australia. B.E. J. Econ. Anal. Policy 7 (2), Article 6.

Leigh, A., 2009. Permanent Income Inequality: Australia, Britain, Germany, and the United States Compared. Discussion Paper 628, Australian National University, Canberra, ACT. http://econrsss.anu.edu.au/pdf/DP628.pdf.

Lerman, J.I., Yitzhaki, S., 1989. Improving the accuracy of estimates of Gini coeffcients. J. Econ. 42, 43–47.

Levine, D., Mazumder, B., 2007. The growing importance of family: evidence from brothers' earnings. Ind. Relat. 46 (1), 7–21.

Lindahl, M., Palme, M., Massih, S.S., Sjögren, A., 2012. Transmission of Human Capital Across Four Generations: Intergenerational Correlations and a Test of the Becker-Tomes Model. Discussion Paper 6463, IZA, Bonn. http://ftp.iza.org/dp6463.pdf.
Long, J., Ferrie, J., 2007. The path to convergence: intergenerational occupational mobility in Britain and the US in three eras. Econ. J. 117 (519), C61–C71.
Long, J., Ferrie, J., 2013a. Intergenerational occupational mobility in Great Britain and the United States since 1850. Am. Econ. Rev. 103 (4), 1109–1137.
Long, J., Ferrie, J., 2013b. Intergenerational occupational mobility in Great Britain and the United States since 1850: reply. Am. Econ. Rev. 103 (5), 2041–2049.
Maasoumi, E., 1998. On mobility. In: Giles, D., Ullah, A. (Eds.), Handbook of Applied Economic Statistics. Marcel Dekker, New York, pp. 119–175.
Maasoumi, E., Trede, M., 2001. Comparing income mobility in Germany and the United States using generalized entropy mobility measures. Rev. Econ. Stat. 83 (3), 551–559.
Maasoumi, E., Zandvakili, S., 1986. A class of generalized measures of mobility with applications. Econ. Lett. 22 (1), 97–102.
Maasoumi, E., Zandvakili, S., 1990. Generalized entropy measures of mobility for different sexes and income levels. J. Econ. 43 (1), 121–133.
Marchon, C.H., 2008. Intergenerational Mobility in Earnings in Brazil Spanning Three Generations and Optimal Investment in Electricity Generation in Texas. Ph.D. thesis, A&M University, Texas. http://hdl.handle.net/1969.1/86018.
Mare, R., 2011. A multigenerational view of inequality. Demography 48 (1), 1–23.
Markandya, A., 1984. The welfare measurement of changes in economic mobility. Economica 51 (204), 457–471.
Mayer, S.E., Lopoo, L.M., 2005. Has the intergenerational transmission of economic status changed? J. Hum. Resour. 40 (1), 170–185.
Mazumder, B., 2005a. The apple falls even closer to the tree than we thought: new and revised estimates of the intergenerational inheritance of earnings. In: Bowles, S., Gintis, H., Groves, M.O. (Eds.), Unequal Chances: Family Background and Economic Success. Princeton University Press, Princeton, NJ, pp. 80–99 (Chapter 2).
Mazumder, B., 2005b. Fortunate sons: new estimates of intergenerational mobility in the United States using social security earnings data. Rev. Econ. Stat. 87 (2), 235–255.
Mazumder, B., 2008. Sibling similarities and economic inequality in the US. J. Popul. Econ. 21 (3), 685–701.
Meghir, C., Pistaferri, L., 2011. Earnings, consumption and lifecycle choices. In: Ashenfelter, O., Card, D. (Eds.), In: Handbook of Labor Economics, vol. 4B. Elsevier, pp. 773–854.
Mendola, D., Busetta, A., 2012. The importance of consecutive spells of poverty: a path-dependent index of longitudinal poverty. Rev. Income Wealth 58 (2), 355–374.
Mendola, D., Busetta, A., Milito, A., 2011. Combining the intensity and sequencing of the poverty experience: a class of longitudinal poverty indices. J. R. Stat. Soc. Ser. A 174 (4), 953–973.
Minicozzi, A.L., 2002. Estimating Intergenerational Earnings Mobility for Daughters. University of Texas at Austin, unpublished manuscript.
Minicozzi, A.L., 2003. Estimation of sons' intergenerational earnings mobility in the presence of censoring. J. Appl. Econ. 18 (3), 291–314.
Mitra, T., Ok, E., 1998. The measurement of income mobility: a partial ordering approach. Econ. Theory 12 (1), 77–102.
Mocetti, S., 2007. Intergenerational earnings mobility in Italy. B.E. J. Econ. Anal. Policy 7 (2), Article 5.
Moffitt, R.A., Gottschalk, P., 2012. Trends in the transitory variance of male earnings: methods and evidence. J. Hum. Resour. 47 (1), 204–236.
Muller, S.M., 2010. Another problem in the estimation of intergenerational income mobility. Econ. Lett. 108 (3), 291–295.
Mulligan, C.B., 1997. Parental Priorities and Economic Inequality. Chicago University Press, Chicago and London.
Naga, R.H.A., 2002. Estimating the intergenerational correlation of incomes: an errors-in-variables framework. Economica 69 (273), 69–92.

Ng, I.Y., Shen, X., Ho, K.W., 2009. Intergenerational earnings mobility in Singapore and the United States. J. Asian Econ. 20 (2), 110–119.
Nicoletti, C., Ermisch, J.F., 2007. Intergenerational earnings mobility: changes across cohorts in Britain. B. E. J. Econ. Anal. Policy 7 (2), Article 9.
Nybom, M., Stuhler, J., 2011. Heterogeneous Income Profiles and Life-Cycle Bias in Intergenerational Mobility Estimation. Discussion Paper 5697, IZA, Bonn. http://ftp.iza.org/dp5697.pdf.
OECD, 1996. Earnings inequality, low paid employment, and earnings mobility. In: Economic Outlook. OECD, Paris, pp. 59–99.
O'Neill, D., Sweetman, O., Van de gaer, D., 2007. The effects of measurement error and omitted variables when using transition matrices to measure intergenerational mobility. J. Econ. Inequal. 5 (2), 159–178.
Österbacka, E., 2001. Family background and economic status in Finland. Scand. J. Econ. 103 (3), 467–484.
Palmisano, F., Van de gaer, D., 2013. History Dependent Growth Incidence: A Characterization and an Application to the Economic Crisis in Italy. SERIES Working Paper 45, University of Bari, Bari. http://ssrn.com/abstract=2231927.
Pekkala, S., Lucas, R.E.B., 2007. Differences across cohorts in Finnish intergenerational income mobility. Ind. Relat. 46 (1), 81–111.
Pekkarinen, T., Uusitalo, R., Kerr, S., 2009. School tracking and intergenerational income mobility: evidence from the Finnish comprehensive school reform. J. Public Econ. 93 (7–8), 965–973.
Piketty, T., 2000. Theories of persistent inequality and intergenerational mobility. In: Atkinson, A.B., Bourguignon, F. (Eds.), In: Handbook of Income Distribution, vol. 1. Elsevier, Amsterdam, pp. 429–476 (Chapter 8).
Porter, C., Quinn, N.N., 2012. Normative Choices and Tradeoffs When Measuring Poverty Over Time. Working Paper 56, OPHI, University of Oxford. http://www.ophi.org.uk/wp-content/uploads/ophi-wp-56.pdf.
Quah, D.T., 1996. Convergence empirics across economies with (some) capital mobility. J. Econ. Growth 1 (1), 95–124.
Raaum, O., Bratsberg, B., Røed, K., Österbacka, E., Eriksson, T., Jäntti, M., Naylor, R., 2007. Marital sorting, household labor supply and intergenerational earnings mobility across countries. B.E. J. Econ. Anal. Policy 7 (2), Article 7.
Rendtel, U., Nordberg, L., Jäntti, M., Hanisch, J., Basic, E., 2004. Report on quality of income data. CHINTEX Working Paper 21, http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Wissenschaftsforum/Chintex/ResearchResults/Einfuehrung,templateId=renderPrint.psml.
Reville, R.T., 1995. Intertemporal and Life Cycle Variation in Measured Intergenerational Earnings Mobility. RAND, Santa Monica, CA, unpublished manuscript.
Riphahn, R., Schnitzlein, D., 2011. Wage Mobility in East and West Germany. Discussion Paper 6246, IZA, Bonn. http://ftp.iza.org/dp6246.pdf.
Rodgers, J.R., Rodgers, J.L., 1993. Chronic poverty in the United States. J. Hum. Resour. 28, 25–54.
Rodgers, J.R., Rodgers, J.L., 2009. Contributions of longitudinal data to poverty measurement in Australia. Econ. Rec. 85, S35–S47.
Roemer, J.E., 2004. Equal opportunity and intergenerational mobility: going beyond intergenerational transition matrices. In: Corak, M. (Ed.), Generational Income Mobility in North America and Europe. Cambridge University Press, Cambridge, pp. 48–57 (Chapter 3).
Rowntree, S., Lavers, G.R., 1951. Poverty and the Welfare State. Longman Green, London.
Ruiz-Castillo, J., 2004. The measurement of structural and exchange income mobility. J. Econ. Inequal. 2, 219–228.
Saez, E., Veall, M.R., 2005. The evolution of high incomes in Northern America: lessons from Canadian evidence. Am. Econ. Rev. 95 (3), 831–849.
Schluter, C., 1998. Statistical inference with mobility indices. Econ. Lett. 59, 157–162.
Schluter, C., Trede, M., 2003. Local versus global assessments of mobility. Int. Econ. Rev. 44, 1313–1335.
Schluter, C., Van de gaer, D., 2011. Upward structural mobility, exchange mobility, and subgroup consistent mobility measurement: US-German mobility rankings revisited. Rev. Income Wealth 57 (1), 1–22.
Schnitzlein, D.D., 2014. How important is the family? Evidence from sibling correlations in permanent earnings in the USA, Germany, and Denmark. J. Popul. Econ. 27 (1), 69–89.

Shin, D., Solon, G., 2011. Trends in men's earnings volatility: what does the Panel Study of Income Dynamics show? J. Public Econ. 95, 973–982.
Shore, S.H., 2011. The intergenerational transmission of income volatility: is riskiness inherited? J. Bus. Econ. Stat. 29 (3), 372–381.
Shorrocks, A.F., 1978a. Income inequality and income mobility. J. Econ. Theory 19 (2), 376–393.
Shorrocks, A.F., 1978b. The measurement of mobility. Econometrica 46 (5), 1013–1024.
Shorrocks, A.F., 1981. Income stability in the United States. In: Klevmarken, N.A., Lybeck, J.A. (Eds.), The Statics and Dynamics of Income. Tieto, Clevedon, Avon, pp. 175–194.
Shorrocks, A.F., 1993. On the Hart masure of income mobility. In: Casson, M., Creedy, J. (Eds.), Industrial Concentration and Economic Inequality: Essays in Honour of Peter Hart. Edward Elgar, Aldershot.
Solon, G., 1989. Biases in the estimation of intergenerational earnings correlations. Rev. Econ. Stat. 71 (1), 172–174.
Solon, G., 1992. Intergenerational income mobility in the United States. Am. Econ. Rev. 82 (3), 393–408.
Solon, G., 1999. Intergenerational mobility in the labor market. In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics, vol. 3. Elsevier Science B.V, New York, pp. 1761–1800.
Solon, G.M., 2002. Cross-country differences in intergenerational earnings mobility. J. Econ. Perspect. 16 (3), 59–66.
Solon, G., 2004. A model of intergenerational mobility variation over time and place. In: Corak, M. (Ed.), Generational Income Mobility in North America and Europe. Cambridge University Press, Cambridge, pp. 38–47 (Chapter 2).
Solon, G., Corcoran, M., Gordon, R., Laren, D., 1991. A longitudinal analysis of sibling correlation in economic status. J. Hum. Resour. 26 (3), 509–534.
Sorensen, J.B., 1992. Locating class cleavages in inter-generational mobility: cross-national commonalities and variations in mobility patterns. Eur. Sociol. Rev. 8 (3), 267–281.
Swift, A., 2006. Political Philosophy: A Beginners' Guide for Students and Politicians, revised ed. Polity Press, Cambridge.
Tawney, R.H., 1964. Equality. George Allen and Unwin, London.
Trede, M., 1998. Making mobility visible: a graphical device. Econ. Lett. 59 (1), 79–82.
Treiman, D.J., 1977. Occupational Prestige in Comparative Perspective. Academic Press, New York.
Ueda, A., 2009. Intergenerational mobility of earnings and income in Japan. B.E. J. Econ. Anal. Policy 9 (1), 1–27.
Van de gaer, D., Schokkaert, E., Martinez, M., 2001. Three meanings of intergenerational mobility. Economica 68 (272), 519–537.
Van Kerm, P., 2004. What lies behind income mobility? Reranking and distributional change in Belgium, Western Germany and the USA. Economica 71 (282), 223–239.
Van Kerm, P., 2006. Comparisons of Income Mobility Profiles. ISER Working Paper 2006-36, ISER, University of Essex, Colchester. http://www.iser.essex.ac.uk/publications/working-papers/iser/2006-36.pdf.
Van Kerm, P., 2009. Income mobility profiles. Econ. Lett. 102 (2), 93–95.
Van Kerm, P., 2011. Picturing mobility: Transition probability color plots. In: Presentation to 2011 London Stata Users Group meeting, London. http://www.stata.com/meeting/uk11/abstracts/UK11_vankerm.pdf.
Van Kerm, P., Pi Alperin, M.N., 2013. Inequality, growth and mobility: the intertemporal distribution of income in European countries 2003–2007. Econ. Model. 35, 931–939.
Xie, Y., Killewald, A., 2013. Intergenerational occupational mobility in Great Britain and the United States since 1850: comment. Am. Econ. Rev. 103 (5), 2003–2020.
Yitzhaki, S., Wodon, Q., 2005. Mobility, inequality, and horizontal inequity. In: Amiel, Y., Bishop, J. (Eds.), In: Research on Economic Inequality, Studies on Economic Well-Being: Essays in Honor of John Formby, vol. 12. Emerald, Bingley, pp. 177–198.
Zandvakili, S., 1992. Generalized entropy measures of long-run inequality and stability among male headed households. Empir. Econ. 17 (4), 565–581.
Ziliak, J.P., Hardy, B., Bollinger, C., 2011. Earnings volatility in America: evidence from matched CPS. Labour Econ. 18 (6), 742–754.
Zimmerman, D.J., 1992. Regression toward mediocrity in economic stature. Am. Econ. Rev. 82 (3), 409–429.

CHAPTER 11

The Global Distribution of Income

Sudhir Anand*, **Paul Segal**†
*Department of Economics, University of Oxford, Oxford, UK; and Department of Global Health and Population, Harvard School of Public Health, Boston, MA, USA
†King's International Development Institute, King's College London, London, UK

Contents

Abstract

This chapter investigates recent advances in our understanding of the global distribution of income, and produces the first estimates of global inequality that take into account data on the incomes of the top one percent within countries. We discuss conceptual and methodological issues – including alternative definitions of the global distribution, the use of household surveys and national accounts data, the use of purchasing power parity exchange rates, and the incorporation of recently available data on top incomes from income tax records. We also review recent attempts to estimate the global distribution of income. Our own estimates combine household survey data with top income data, and we analyze various aspects of this distribution, including its within- and between-country components, and changes in relative versus absolute global inequality. Finally, we examine global poverty, which is identified through the lower end of the global distribution.

Handbook of Income Distribution, Volume 2A
ISSN 1574-0056, http://dx.doi.org/10.1016/B978-0-444-59428-0.00012-6

Keywords

Global inequality, purchasing power parity exchange rates, household surveys, national accounts, top incomes, global poverty

JEL Classification Codes

D63, E01, I32

11.1. INTRODUCTION

As the world has become increasingly interconnected through trade, investment, migration and communication, people's interest in and knowledge of international comparisons of living standards has grown. Correspondingly, the global distribution of income has become the subject of numerous research papers and articles, and commentaries in the media. In the popular imagination it seems self-evident to be of interest that great wealth and great poverty coexist in the world. In this chapter we examine the concept of global inequality, the normative motivations for studying it, and the available evidence on the global distribution of income. Widely varying estimates of global income inequality have been published, using a variety of data and methodologies. We critically discuss the different approaches and assumptions behind them, with a view to determining what we believe to be best practice. We also construct a global distribution using both household surveys and top income shares from tax data.

Inequality is a broad concept, and the global distribution of income allows various interpretations. For this reason we start by clarifying different conceptions of the global distribution of income. The distribution of primary interest for us, and the subject of most of this chapter, is that among individuals in the world, each assigned his or her per capita household income. This is what we will refer to as the global distribution of income. But other distributions of global income are also of interest for certain questions. Studies of economic growth and convergence, for instance, are based on changes in the distribution among countries of per capita national income, which is a type of global income distribution that is only indirectly related to the global distribution of income among individuals.

Because individuals around the world are naturally partitioned by country of residence, we examine the between-country and within-country components of global inequality, which can have different definitions depending on the inequality measure used. Although we do not discuss the causes of changes in global income inequality, this decomposition provides a breakdown of those changes, allowing us to isolate the contributions of differential growth in per capita income across countries, and of changes in inequality within countries. This decomposition is a necessary precursor to any causal explanation because one would expect different mechanisms to explain the two components.

Studying the global distribution of income raises difficult empirical and measurement issues. To compare real incomes across countries one needs to convert them using purchasing power parity (PPP) exchange rates, rather than market exchange rates, to take account of aggregate price differences between countries. There are different methods for calculating PPP exchange rates, which have their respective merits and are used by different studies. All methods depend on the price surveys conducted by the International Comparison Program (ICP). In some cases those price surveys have themselves been controversial. We do not discuss PPP exchange rates in detail (q.v. Anand and Segal, 2008), but we highlight some of the features and controversies that are most relevant for studying the global distribution of income.

Another empirical controversy concerns the measurement of mean incomes within countries. Any global distribution of income must rely on national household surveys to estimate inequality within countries. But some studies, instead of using the mean incomes recorded by those surveys, have taken the relative distributions implied by them and "scaled" them to national accounts estimates of per capita GDP or household consumption expenditure. We argue that there is no good reason to scale to GDP, but that the use of household final consumption expenditure (HFCE) from the national accounts, which is available for most countries, may provide a useful robustness check. Using HFCE rather than the mean incomes from household surveys changes both the level and trend of estimated global inequality.

Beyond reviewing the conceptual and measurement issues underlying the study of the global distribution of income, the empirical aim of this chapter is to use the best available data to construct global distributions of income based on alternative plausible assumptions. The main innovation is to supplement data from household surveys with newly available estimates of top income shares derived from tax data in a range of countries. These data constitute a significant advance in our understanding of the distribution of income both within countries and globally because individuals at the top of the income distribution are either not represented or are underrepresented in household surveys. Unsurprisingly, their inclusion leads to substantially higher estimates of global inequality.

The chapter continues as follows. Section 11.2 discusses the motivation for the study of global income inequality. Section 11.3 analyzes the different concepts of the global distribution of income. Section 11.4 discusses methodological issues and describes the available data, including the top income share data. Section 11.5 presents our constructed global distributions of income and the corresponding estimates of global inequality. Section 11.6 decomposes global income inequality into between-country and within-country inequality and discusses their significance and evolution. Section 11.7 examines the distinction between relative and absolute inequality and presents some preliminary estimates of absolute global inequality. Section 11.8 turns to the estimation of global poverty and considers its level, trends, and regional concentration. Section 11.9 concludes the chapter.

11.2. WHY STUDY THE GLOBAL DISTRIBUTION OF INCOME?

Interest in global inequality reaches far beyond academia and has increased dramatically in recent years—among activists and NGOs, the news media, and national and international institutions and policymakers. This is in part due to the perception that the benefits of rapid economic growth in recent decades, which has coincided with a period of rapid globalization, have been distributed highly unequally. Thus, the worldwide "Occupy" movement launched in 2011, with its slogan "We are the 99%," has focused on the sharply increasing concentration of income and wealth among the top 1% of income recipients compared to the other 99%. In the news media, *The Economist* has described growing inequality as "one of the biggest social, economic and political challenges of our time" (Beddoes, 2012). At the 2012 World Economic Forum meeting at Davos, "severe income disparity" was featured as the single most likely global risk, and with one of the highest potential impacts.[1] Again at Davos in 2013, Christine Lagarde, managing director of the International Monetary Fund, stated that "[e]xcessive inequality is corrosive to growth; it is corrosive to society. I believe that the economics profession and the policy community have downplayed inequality for too long" (Lagarde, 2013).

There is indeed a positive case for being concerned about the consequences of inequality for economic growth and social cohesion; crime rates and population health, for instance, have been linked to income inequality within countries.[2] To the extent that such (within-) country inequality contributes to global inequality, there will be a corresponding concern about the latter. One might equally be concerned about the "corrosive" effects of global inequality itself. Davos, where Lagarde made her comments, is a meeting place of the global élite (i.e., those at the top of the global income distribution, and not just their respective national distributions).

The normative case for studying global inequality seems obvious to some, but it is contested by philosophers who believe that the distribution of income among individuals can be a matter of justice only if they share a government. Nevertheless, even these philosophers typically agree that "there is some minimal concern we owe to fellow human beings threatened with starvation or severe malnutrition and early death from easily preventable disease," and that therefore "the urgent current issue is what can be done in the world economy to reduce extreme global poverty" (Nagel, 2005, p. 118). In itself this warrants study of at least the lower end of the global distribution of income.

An alternative understanding of justice may lead to a normative concern about global inequality. Some cosmopolitan political theorists argue that egalitarian principles apply equally at the global as at the national level simply because all human beings are entitled to equal respect and concern.[3] On this view, national borders are not relevant to an ethical concern with inequality.

It can also be argued that the institutional arrangements that exist in the global economy—the international rules and organizations that govern the flows of goods,

[1] World Economic Forum (2012), reported by Tett (2012).
[2] See, for example, Pickett and Wilkinson (2010).
[3] For discussion see Sen (2000) and Bernstein (2011).

capital, and labor between countries—are sufficient to generate a normative concern for inequality among individuals in the world, even if they fall short of a global government. These international arrangements are largely determined by rich countries and tend to benefit citizens in rich countries at the expense of citizens in poor countries. Rich countries may therefore bear some responsibility for global inequality. Sen (2009, p. 409) puts the issue as follows: "The distribution of the benefits of global relations depends not only on domestic policies, but also on a variety of international social arrangements, including trade agreements, patent laws, global health initiatives, international educational provisions, facilities for technological dissemination, ecological and environmental restraint, treatment of accumulated debts (often incurred by irresponsible military rulers of the past), and the restraining of conflicts and local wars."

In studying the global distribution of income, we need to distinguish between the recognition of inequality and the obligation and capacity to reduce it. Through its domestic policies, a sovereign state can have more influence on national inequality than on global inequality. This might suggest that, from a policy viewpoint, we should assess within-country inequality differently from between-country inequality (see Section 11.6)—especially if international institutions have limited powers to address between-country inequality. In any case, as we improve our understanding of global inequality, we will be in a better position to diagnose its causes and discuss ways of mitigating it.

In this chapter we take the global distribution of income to be of intrinsic interest. We will analyze various aspects of this distribution, including its within- and between-country components, and also its lower end, which is needed to identify global poverty. Constructing the global distribution of income can be the first step in a broad exercise, which should ultimately permit us to examine many different aspects of global inequality—such as the extent to which gender, ethnicity, education and other socioeconomic variables contribute to global inequality, the characteristics of the global poor, and the composition of the global top 1%.

11.3. WHICH GLOBAL DISTRIBUTION OF INCOME?

Our starting point must be to clarify what we mean by the global distribution of income. Following Milanovic (2005) and Anand and Segal (2008), we can define four concepts of the global distribution of income and their associated levels of inequality, distinguished by the population unit and the income concept (which may be a measure of consumption expenditure) to which they refer. The four concepts of global distribution are relevant to addressing quite different questions, as we discuss in this section. We must also decide on the numéraire to make the income concept comparable across countries, and use either market exchange rates or PPP exchange rates. PPP exchange rates account for the fact that one U.S. dollar will typically buy less in the United States than one dollar's worth of, for example, Indian rupees purchased on the currency markets, will buy in India. Later we will discuss different approaches to PPP exchange rates and some of the complications that arise in estimating and using them. Which exchange rate is appropriate will depend on the question being asked.

Our first concept of the global distribution of income, denoted *concept 0*, is the distribution of global income by country. In other words, the "population unit" is the *country* and the "income concept" is the (total) *national income* of the country. Thus India and Canada, both with GDPs of US$1.8 trillion in 2012, count as equal, despite the fact that India has a population of 1237 million and Canada only 35 million. It is this concept 0 global distribution that is most relevant for questions of geopolitics and market access. In international negotiations over trade rules and macroeconomic policies it is a country's total economic size that tends to determine its bargaining power. For such questions, it is a country's weight in international markets—its command over internationally traded goods and services, or financial assets—that matters, and hence income at market exchange rates is likely to be relevant. One might refine the measure, of course, depending on the geopolitical question at hand. For example, in matters concerning global energy markets, countries with relatively small economies but large fuel exports tend to be important.

Next, the *concept 1* global distribution again takes the *country* as the population unit, but now the income concept is the *national income per capita* of the country, not its total national income.[4] This is the concept typically used in analyses of economic growth, and in particular of economic convergence, where the question is how the set of characteristics and policies associated with a given country affects its per capita income growth rate. Because it is real output that is of interest in this case, income levels will be measured at PPP exchange rates.

In the *concept 2* global distribution, the population unit is the *individual*, and the income concept is again *national (household) income per capita*. (This is equivalent to taking the country as the population unit, as in concept 1, but weighting each country by the size of its population.) It is not obvious why this concept of global inequality would be intrinsically interesting, but some older studies have analyzed its evolution over time, mainly because of the ready availability of data on national income or GDP per capita (Boltho and Toniolo, 1999; Firebaugh, 1999, 2003; Melchior et al., 2000). Concept 2 is of instrumental interest, however, through its relationship with concept 3, which is the focus of this chapter.

The *concept 3* distribution also takes the *individual* as the population unit, but the income concept is the *per capita income of the household to which the individual belongs*—under the assumption of equal sharing of household income (or consumption expenditure). It is the global analogue of the type of distribution typically used to calculate inequality within

[4] It follows that this *concept 1* distribution—unlike the concept 0 distribution—is not a "distribution of (total) global income among countries."

Table 11.1 Concepts of global income distribution and inequality

	Unit of analysis	Ranking variable	Numéraire
Concept 0	Country	National income	US$ or PPP$
Concept 1	Country	National income per capita	PPP$
Concept 2	Individual	National (household) income per capita	PPP$
Concept 3	Individual	Household income per capita of individual	PPP$

countries.[5] Henceforth we use the terms "global distribution of income" and "global inequality" without further qualification to refer to their *concept 3* counterparts. Because it is real income or consumption that we are interested in, national currencies will be compared using PPP exchange rates. Concept 3 is also the only concept that tells us something directly about global welfare.

Concept 2 global inequality can be seen as the between-country component of concept 3 global inequality. Concept 2 inequality tells us what concept 3 inequality would be if there were no inequality within countries and each person in a country received the national (household) income per capita of that country. For decomposable measures of inequality, concept 3 inequality will then be equal to concept 2 inequality plus a weighted average of inequality within countries (the within-country component of concept 3 inequality). We will discuss these distinctions further when we present our calculations later in this chapter.

Table 11.1 summarizes the four global income distributions defined in terms of unit of analysis (population unit), associated ranking variable (income concept), and numéraire. It is important to emphasize that the four different concepts of global inequality can move in different directions. It should be immediate from the decomposition just mentioned that concepts 2 and 3 can move in different directions: a modest fall in between-country (i.e., concept 2) inequality may coexist with a rise in concept 3 global inequality if within-country inequality increases sufficiently.

Moreover, the same changes in national income may have opposite effects on different concepts of inequality. For example, China is the second-largest economy in the world, both in PPP$ and in current US$, but its per capita GDP in 2012 was PPP$7960, below both the unweighted mean GDP per capita across countries of PPP$12,300 and the

[5] In country studies of income inequality, an adjustment for differential needs and economies of scale in household consumption is sometimes made by taking account of the age and sex composition of a household, in addition to its size. This is done through "equivalence scales," which allow the calculation of the number of "equivalent adults" in a household. Each individual in the household is then assigned the household's income per equivalent adult. Given the type of survey data at our disposal, it is not possible to estimate the number of equivalent adults in each household and rank individuals by their household income per equivalent adult. Hence, like other studies of interpersonal global inequality, we simply rank individuals by their household income per capita.

population-weighted mean across countries of PPP$10,260.[6] The fact that China's above-average total national income has been growing much faster than the world average therefore implies that China is a disequalizing force for concept 0 inequality. However, the fact that its below-average GDP per capita is growing faster than the world means of both unweighted and population-weighted GDP per capita is an equalizing force for concepts 1 and 2 global inequality, respectively. The latter implies that it is also an equalizing force for concept 3 global inequality.

Consider now the notion of "convergence" in the literature on economic growth, which is the closest that many economists get to thinking about global inequality. There are two commonly used definitions of "convergence," namely beta and sigma convergence. Beta convergence means that when a country's growth rate is regressed on its national income per capita, the coefficient on income is negative and significant.[7] Thus, on average, countries with higher per capita national income (where per capita GDP is the measure typically used in these studies) have lower growth rates. Sigma convergence means that the dispersion across countries of per capita national income declines over time, often measured by the standard deviation of the logarithm of per capita national income. Both therefore refer to the concept 1 global distribution with the country as the population unit and per capita national income as the income concept.

In their survey of growth econometrics, Durlauf et al. (2009, p. 1098) state that sigma convergence has "a natural connection to debates on whether inequality across countries is widening or diminishing." If "inequality across countries" refers to concept 1 inequality, then it is a tautology that sigma convergence will measure "whether inequality across countries is widening or diminishing."[8] However, sigma convergence or divergence has no necessary connection to any other concept of inequality across countries (e.g., concept 0 inequality as seen in the China example) or to global inequality across a different population unit (e.g., individuals). A rise in the dispersion across countries of per capita national income (i.e., concept 1 inequality) may be associated with a fall in concepts 2 and 3 global inequality, as the following example demonstrates.

The Philippines has a population of 97 million people, and a per capita GDP of PPP$3800. Its per capita GDP is below both the unweighted and the population-weighted world means of PPP$12,300 and PPP$10,260, respectively, noted earlier. There are 35 countries that both have populations below 5 million and also, like the Philippines, have per capita GDP below the respective unweighted and population-weighted world means. For the purposes of sigma convergence, each of these 35 small

[6] PPP$ are in 2005 prices, and data are from *World Development Indicators*, http://data.worldbank.org/data-catalog/world-development-indicators.

[7] *Conditional* beta convergence means that the coefficient on income is negative and significant when other variables are controlled for in the regression.

[8] One concern about this argument is that the standard deviation of log-income is not a good inequality measure, as it does not satisfy the principle of transfers at the top end of the income distribution (Sen, 1973).

countries has the same weight as the Philippines, yet their combined populations amount to 57 million, below that of the Philippines.[9] Now imagine that there is sigma divergence, where all other countries are growing at a common rate, but the Philippines is growing faster and the 35 small countries are growing slower than this common rate. Global inequality is increasing according to concept 1 because while one country (the Philippines) whose per capita GDP is below the unweighted world mean is converging to the world mean, 35 other countries whose per capita GDP is below the world mean are diverging from it. But global inequality may be decreasing according to concept 2 because the convergence of the Philippines' large population toward the weighted world mean outweighs the divergence of the populations of the 35 small countries away from the weighted world mean. Assuming that inequality within countries is unchanged, global inequality may therefore also be decreasing according to concept 3.

We conclude that global inequality *tout court* is an underspecified concept, and estimates of different definitions of global inequality can move in different directions—as we find in our empirical estimates in Sections 11.5 and 11.6.

11.4. DATA

11.4.1 Household Surveys and National Accounts

Household surveys are the most widely available source of data for estimating income distributions within countries, and it is the great expansion in their global coverage that has permitted estimates of global inequality. One could in principle use census data or other sources—but in practice these are available for far fewer country-years than household surveys. Survey coverage has expanded dramatically in the last 30 years; the surveys used by the World Bank to estimate global poverty in 1981 covered only 51.3% of the population of the developing world, whereas in 2005 they covered 90.6% (Chen and Ravallion, 2008).

Although there is no credible alternative to using household surveys for estimates of global inequality, they do suffer limitations. Beyond the obvious sampling and measurement errors, surveys may suffer from biases due to underreporting of incomes by the rich and undersampling of both very rich and very poor households. Most important for our purposes, differences in definitions and coverage mean that different surveys are typically not strictly comparable with one another (see Anand and Kanbur, 1993, pp. 33–36). Atkinson and Brandolini (2001) described such problems in the Deininger and Squire database, which collates estimates of inequality within countries; Anand and Segal (2008) discussed these issues in the context of measuring global inequality, observing that in some surveys incomes are gross-of-tax and in others net-of-tax; some refer to cash

[9] GDP and population data are from *World Development Indicators*, http://data.worldbank.org/data-catalog/world-development-indicators.

incomes, whereas others include certain items of income-in-kind; some impute the rental value of owner-occupied housing, whereas others do not. Moreover, all global data sets of household surveys combine surveys of income and of consumption expenditure. There is no reliable way to infer an income distribution from an expenditure distribution, or vice versa, so one simply has to live with the noncomparability. For brevity we will refer to "income or consumption expenditure distributions" as "income distributions."

The World Bank's Living Standards Measurement Surveys, initiated in 1980, have been instrumental in increasing both the quantity of survey data available and its quality. The Luxembourg Income Study (LIS) specifically attempts to harmonize survey data to ensure their comparability, and the LIS data set currently covers 47 countries. Still, non-comparability cannot be avoided in a global data set of household surveys, which cover most of the world's population.

Although all recent studies of global inequality use survey data for estimates of within-country inequality, most then "scale" the within-country distributions to national accounts estimates of mean income or consumption expenditure. For instance, Chotikapanich et al. (1997), Dowrick and Akmal (2005), Sala-i-Martín (2006), and Schultz (1998) use the Deininger and Squire (1996) inequality database for estimates of *relative* inequality within countries and peg the relative distributions around an absolute mean from the national accounts.[10] Milanovic (2002, 2005, 2012) and Lakner and Milanovic (2013) are the only studies we know of that estimate global income inequality using levels of income or expenditure directly from surveys, rather than scaling relative distributions to NA means (though Lakner and Milanovic do use NA means in imputing top incomes, as we discuss later). The World Bank also uses absolute incomes from household surveys for its estimates of global poverty (Chen and Ravallion, 2001, 2008, 2012).[11] The distinction between using survey data directly and scaling them to national accounts categories matters because both the levels and rates of change of global inequality and poverty can vary substantially (Deaton, 2005).

For studies that use household surveys only for their relative distributions and scale them to national accounts means, there are two widely available estimates of NA "mean income": per capita GDP and per capita household final consumption expenditure (HFCE). In principle, one would want to use the category of personal income, but countries do not usually report this category. Most studies of global inequality simply use per capita GDP as a proxy for individual mean (per capita household) income.[12]

[10] See Anand and Segal (2008) for detailed descriptions of their methodologies.

[11] The World Bank's own estimates are based on unit record data. These data are available to the public only in coarser, grouped form from the World Bank's Povcalnet Website: http://iresearch.worldbank.org/PovcalNet/index.htm.

[12] These studies are Bourguignon and Morrisson (2002), Bourguignon (2011), Sala-i-Martín (2006), Dowrick and Akmal (2005), Schultz (1998), Chotikapanich et al. (1997), and Korzeniewicz and Moran (1997). Dikhanov and Ward (2002) use "personal consumption expenditure," which we take to mean HFCE.

As argued in Anand and Segal (2008, pp. 66–68), if it is national household consumption expenditure that one wishes to measure, then there is no reason to use GDP when HFCE is available. Moreover, GDP is also a poor measure of *household income*: GDP includes depreciation, retained earnings of corporations, and the part of government revenue (taxes) that is not distributed back to households as cash transfers. Deaton (2005, p. 4) noted that "much of saving may not be done by households, but by corporations, government, or foreigners, so that household income may be closer to household consumption than to national income." In the case of the United States, which is one of the few countries that does report measures of aggregate household income (referred to as "personal income"), it amounts to only about 70% of GDP. Deaton estimates that, across 272 surveys of household income from around the world, survey household income amounts on average to only 57% of GDP, but equals 90% (101% population-weighted) of HFCE from National Accounts.

The question remains, however, whether one would want to use *any* National Accounts figures when mean household income (or consumption) is available in the surveys themselves, which are the source of the income (or consumption) distribution for countries. We saw earlier that surveys have their own problems. But they are at least a direct measure of the variable of interest. HFCE, on the other hand, includes the category of "non-profit institutions serving households" (e.g., religious organizations and political parties), and suffers from being calculated as a residual of aggregate consumption minus estimates of firms' consumption and government consumption. Errors in any of the latter magnitudes will translate into errors in estimates of HFCE.[13]

New evidence on national accounts data in low-income countries casts more general doubts on their reliability. Jerven (2013) noted that Ghana revised its GDP upward by 60.3% in November 2010 owing to a change in base year,[14] and argues that similarly large revisions are to be expected in other sub-Saharan African countries.[15] Young (2012) also found that national accounts provide a poor measure of growth in sub-Saharan Africa and produces independent estimates of consumption growth based on data from Demographic and Health Surveys.[16]

Most of our analysis that follows will refer to the global distribution based on mean incomes from household surveys, but we also calculate global inequality where mean per capita household income is taken to be equal to per capita HFCE as reported in the National Accounts and compare the differences in the results.

[13] See Anand and Segal (2008, p. 68) for detailed discussion.

[14] The 1993 base-year estimates excluded parts of the economy that were important in the new base year of 2006 (Jerven, 2013, p. 11).

[15] These countries are Nigeria, Uganda, Tanzania, Kenya, Malawi, and Zambia. Jerven's explanation is that many of these countries suffered drastic cuts to statistical services in the 1980s and 1990s.

[16] Note, however, that Young's method of inferring aggregate consumption from data on assets has been criticized by Harttgen et al. (2013).

11.4.2 Top Income Data

Perhaps the most important recent innovation in estimating national income inequality has been the collation of data on top income shares from income tax records. These estimates present the incomes of the top 0.1%, top 1%, and top 10% as a share of "control" income, where control income is an estimate of total personal income in the economy (not just taxable income). They are important primarily because they make a substantial difference to estimated inequality. Household surveys typically undersample (exclude) the richest individuals or underreport their incomes, or both. In the United States in 2006, for instance, tax data excluding capital gains imply a top percentile share of 18.0%, whereas survey data imply a share of 13.7%. Using data for 2006, the U.S. Gini based on household survey data (the Current Population Survey) is 0.470, whereas correcting the top percentile's income using the tax data raises it by nearly 0.05 to 0.519. Moreover, the increase in the U.S. Gini from 1976 to 2006 using survey data alone (corrected for a change in definition) was 0.053, which more than doubles to an increase of 0.108 using the top income data (including capital gains; Atkinson et al., 2011, p. 31; see Burkhauser et al., 2009 for further discussion of U.S. data).

Atkinson et al. (2011, pp. 4–5) describe the top income data in detail, and discuss their limitations. These include the fact that the income shares refer to gross income before tax; the data vary with respect to the unit of observation, some referring to individuals and others to households; in some cases they are not consistent over time, as tax regimes change; and they may be biased owing to tax avoidance and tax evasion. Although they are typically much better than surveys at capturing capital income, this varies depending on the extent to which capital income is taxed and hence reported in the tax records (Atkinson et al., 2011, p. 35). Alvaredo and Londoño Vélez's (2013) study of top incomes in Colombia notes that different definitions of the control income, of which top incomes are expressed as a share, lead to somewhat different estimates. For these reasons international comparisons of these top income shares may suffer from inconsistencies. Nonetheless, we will set aside such concerns and use these data on the presumption that excluding them would cause a large negative bias in estimates of global inequality. Clearly, however, these noncomparabilities do add uncertainty to the estimates.

11.4.3 PPP Exchange Rates

International comparisons of living standards require the use of PPP exchange rates to convert national currencies into a common numéraire.[17] Two standard sets of PPPs are publicly available: those produced by the International Comparison Program (ICP) of the World Bank (World Bank, 2008) and those produced by the Penn World Tables (PWT), which also uses the underlying price survey data collected by the 2005

[17] An early discussion of this issue may be found in Berry et al. (1983).

ICP.[18] PPPs for years before and after the "benchmark" year 2005 are derived from each country's domestic price indices.

The price surveys undertaken for the 2005 ICP were both more detailed and more representative globally than in previous rounds of the ICP. China had never taken part in an ICP before the 2005 round, and India had not taken part since 1985, but both countries were surveyed in the 2005 ICP. Previous estimates of PPPs were therefore based on imputations. Partly for this reason, the results from the 2005 ICP have in some cases led to dramatic changes in estimated GDP. Both China and India were found to have real GDPs nearly 40% lower than previous estimates[19] because prices were found to be higher than previously estimated. In the case of China, at least some of this downward revision appears to have been due to sampling problems: its price surveys took place in cities and their environs and did not cover rural areas. For this reason Chinese prices are likely to have been overestimated, and its real income underestimated. Following Chen and Ravallion (2010), and like Milanovic (2012), we make an adjustment to account for this (described later). Milanovic (2012) found that the revisions in the 2005 ICP make a substantial difference to estimated global inequality, raising the Gini by 4.4–6.1 percentage points over the period 1988–2002 and Theil T by 12.5–16.4 percentage points. Other studies that use the 2005 PPPs are Lakner and Milanovic (2013) and Bourguignon (2011), and we discuss their findings later.

Starting from the vector of prices in each country provided by the ICP, the World Bank and PWT use different methods to calculate PPPs. World Bank PPPs are based on the Elteto–Köves–Szulc (EKS) method, whereas PWT uses the Geary–Khamis (GK) method (both with a variety of adjustments made in the process of estimation).[20] EKS arose from a statistical approach to index numbers (Deaton and Heston, 2010) and is a multilateral generalization of the Fisher index for two countries (for further discussion, see Anand and Segal, 2008, p. 71). However, under certain assumptions EKS applied to incomes yields an index of real living standards, or utility, and for this reason Neary (2004) included it as an example of the "economic" approach to index numbers. Under the economic approach it is assumed that observed quantities arise from the optimizing behavior of some representative agent with a well-defined utility function. Real relative

[18] A new ICP with base year 2011 was released recently in June 2014, as this chapter was already in press.

[19] This was calculated by comparing the countries' respective incomes relative to U.S. income in 2005 at 1993 PPP$ and at 2005 PPP$.

[20] See Anand and Segal (2008) for details of these two methods, and for discussion of the "Afriat method" used by Dowrick and Akmal (2005) to measure global inequality. The World Bank's PPPs use EKS within regions of countries and then link regions using a "ring" of 18 countries with at least two in each region. See Deaton and Heston (2010) for discussion of both the World Bank and PWT PPP methods. These authors also noted that a single global EKS calculation leads to some nontrivial differences compared to the ICP PPPs, including a real GDP in China that is 6.6% higher.

incomes measured using EKS PPPs represent relative utility levels when utility is quadratic (i.e., in these circumstances it is a "true" index).

GK, on the other hand, is an example of the "test" or "axiomatic" approach. The GK index has no interpretation in terms of optimizing behavior, but its putative advantage with respect to EKS is that it passes the test, or obeys the axiom, of *matrix consistency*. That is to say, GK provides a vector of "international prices" for individual goods that enable disaggregation of the economy into subsectors whose values at those prices sum to the total value of the economy. This is not true of EKS, which computes the relative size of aggregate incomes but does not provide a set of international prices with which economies can be consistently disaggregated. If one is interested in analyzing the structure of economies, then matrix consistency would seem to be a useful property. For instance, it is hard to interpret the relative size of manufacturing in two different countries when manufacturing plus nonmanufacturing within each country does not add upto 100% of its economy.

Matrix consistency would seem less relevant, however, when our concern is international comparisons of living standards. In this case, it is the overall value of consumption, not its composition, that concerns us. More important for our purposes is the drawback of the GK method, which is that it suffers from Gershenkron (or substitution) bias. Because consumers tend to substitute away from goods that are relatively expensive and toward goods that are relatively cheap, valuing the output of both country A and country B at country B's prices will lead to an overestimation of the income of country A relative to that of country B. The relative prices arising from the standard GK method more closely resemble those in rich countries than in poor countries, leading to an overvaluation of the incomes of poor countries relative to rich countries and therefore to an underestimation of inequality between countries. Ackland et al. (2004) found that the GK method overvalues the incomes of poorer countries compared to EKS. They regress log per capita GDP from GK on log per capita GDP from EKS and find the slope to be 0.94 and to be significantly less than 1.0. Deaton and Heston (2010) found that the Gini for concept 2 (between-country) global inequality, with per capita GDP as the income concept, is slightly higher using EKS than GK, at 0.533 as opposed to 0.527.

Almås (2012) also found that PWT PPPs underestimate global inequality when accounting for both substitution bias and differences in the quality of goods across countries. However, her estimates are based on the strong assumption that "there is a stable relationship between the budget share for food and household income; i.e., there is a unique Engel relationship for food in the world" (Almås, 2012, p. 1094). Deaton and Heston (2010, p. 5) pointed out that "there are many places in the world, such as North and South India, where there are large differences in consumption patterns of food in spite of only modest differences in relative prices."

Neary (2004) presents a method that he denotes "Geary–Allen International Accounts" (GAIA) for constructing PPPs that is "economic" in the sense of being based on the assumption of optimizing behavior and therefore does not suffer from substitution bias, but that also satisfies a form of matrix consistency. However, the form of matrix consistency satisfied is not the form that GK satisfies; the sectoral quantities that sum to the value of the whole economy are not the actual observed sectoral quantities, but virtual quantities that a reference consumer, whose preferences are estimated from the data, *would* have chosen. So it is also the case in the GAIA method that observed manufacturing plus observed nonmanufacturing within an economy will not, in general, add upto 100% of the economy.

The theoretical advantage of GAIA over EKS is that it is a "true" index (i.e., produces estimates of relative real incomes that are consistent with optimizing behavior) for a wider range of utility functions. But because all such indices make the false assumption of identical tastes in all countries worldwide, this seems a rather limited benefit. EKS, on the other hand, has the advantage of being relatively transparent. Although GAIA requires the estimation of a demand system, the EKS exchange rate for a country is simply the geometric mean of that country's Fisher price indices relative to every other country and, as already mentioned, has a natural statistical interpretation that is attractive to national income accountants if not to consumer theorists (Deaton and Heston, 2010).

In our calculations that follow, we use the EKS-based World Bank consumption PPPs from the 2005 ICP. Following Chen and Ravallion (2008, 2010) we make the following adjustments. For both India and China, where the survey data are provided separately for rural and urban strata, we deflate urban incomes relative to rural incomes from price indices used for the construction of domestic urban and rural poverty lines. For India we assume that the World Bank estimated PPP is a weighted average of the urban and rural PPPs. For China we assume that the reported PPP is for urban areas and adjust rural prices downward. This is because the price surveys in China in 2005 were restricted to 11 metropolitan areas, which did not include any rural areas (Chen and Ravallion, 2010). The result is a lower overall price level for China, and thus higher average living standards, than those implied by the use of the 2005 ICP.

A limitation to all standard PPP estimates is that they assume all households within a country face the same price level for their expenditure basket. This may be problematic for at least two reasons. First, urban and rural areas typically have different price levels, and although we have taken this into account for China and India, where the urban and rural price surveys are distinct, it is not possible to do so for most countries. Second, different quantiles of a national income distribution will typically consume different baskets of goods and services,[21] and hence face different costs of living. For instance, the poor may face higher unit costs for a good because they have to buy it in smaller quantities. Moreover, they purchase goods in different proportions from the nonpoor so the prices of goods will have different expenditure weights for them. At the other end of the

[21] Deaton and Dupriez (2011) discussed this in the case of the poor and have come up with PPPs specifically for estimating global poverty.

distribution, the very rich (such as those captured by the top income data) may tend to buy more goods from outside their country of residence, to which market exchange rates would apply. But to the extent that the very rich spend their income on nontradable goods and services—for example, country estates, urban mansions, and domestic labor within their country of residence—PPPs with different expenditure weights may be more appropriate than market exchange rates.

11.4.4 Estimation Errors

The preceding discussion of the available data indicates that there are several sources of error in estimates of global inequality, including our own. These include sampling errors, which arise from the sample not being representative of the world population. Our global income distribution is constructed as the union of national income distributions, each of which is based on a national household income (or expenditure) survey with a distinct sampling frame and sampling errors (including undersampling of both the rich and the poor in a country). This global distribution is not estimated from a stratified random sample of the world population, so standard methods are not applicable to calculate sampling errors or confidence intervals for estimates of global inequality.

It is important to distinguish sampling errors from other types of estimation error, which arise from imprecise data and invalid or inaccurate assumptions and methods used to calculate global inequality. For example, there are measurement errors in the income or expenditure data in household surveys (e.g., underreporting of incomes of the rich) and in any national accounts data that may be used; there are also estimation errors in the PPP exchange rates used to construct a global income distribution from national distributions. Major revisions in the estimation of PPPs in the 2005 ICP round, discussed earlier, suggest great sensitivity to the assumptions and methods employed. Given such instability, we may expect further revisions in the next set of PPPs from the 2011 round of the ICP.[22] Moreover, as mentioned earlier, a single PPP exchange rate for a country may fail to capture differences in price levels faced by households in different quantiles of the income distribution or in different geographical locations in the country.

Bourguignon and Morrisson (2002) estimated global inequality from 1820 to 1992 through the use of inevitably limited data and manifold assumptions. Given the limitations of their data, they simulated "uncertainty" in their mean income (i.e., GDP) numbers and in their country-group distributions (11 data-points for each of 33 countries or groups of countries) and calculated standard errors for global inequality on this basis. Under their simulation assumptions, the resulting standard errors on the global Gini turn out to be small: in 1820 the standard error is 0.9 Gini points, in 1950 it is 0.2 Gini points, and in 1992 it is 0.1 Gini points (where 1 Gini point is 0.01 in the Gini scale of

[22] These PPPs were not available at the time of writing.

0.00–1.00). In our view the other sources of error discussed earlier would imply much larger confidence intervals than these standard errors suggest.

11.5. ESTIMATING THE GLOBAL DISTRIBUTION OF INCOME

In this section we present new estimates of the global distribution of income that combine household survey data with top income data. These estimates are constructed from Milanovic's (2012) global distribution data set of household surveys for five "benchmark" years in the period 1988–2005, which we have supplemented with top income estimates from income tax data. Milanovic's data are provided in quantiles, in most cases 20 income groups each comprising 5% of the population. For those countries for which Milanovic (2012, pp. 10–11) has unit record data, he compared inequality based on individual records with that based on the constructed vigintile (5%) shares and found that the underestimation of the Gini using vigintile shares varied from 0.001 to 0.006 with a mean of 0.003. We agree with Milanovic that this is small enough to be inconsequential.

The five benchmark years 1988, 1993, 1998, 2002, and 2005 each have surveys for between 103 and 124 countries and cover between 87% and 92% of the world population and between 95% and 98% of global GDP in PPP$. The Milanovic data set provides incomes in national currencies, which we convert to our numéraire of international dollars using World Bank PPPs.[23] We thus have incomes from household surveys in PPP$ for 87% of the world population in 1988, and 90–92% in the later years. There are 67 countries for which we have both survey and PPP data in *all* five benchmark years, which we refer to as the "common sample over time".

As seen in Table 11.2, we have a total of 537 country-years in our data set. Of these, 104 country-years, ranging from 18 to 23 countries in each year, also have income tax data on the share of the top percentile of the population, which we downloaded from the World Top Incomes Database.[24] These countries include the three largest developing countries—China, India, and Indonesia; one Latin American country—Argentina; one African country—South Africa; and all the G7 countries.

The rationale for using income tax data for top percentile shares is that household surveys typically fail to capture the incomes of the richest members of society. For example, Székely and Hilgert (1999) found that in most surveys in Latin America the richest

[23] For Soviet republics in 1988 we use Milanovic's calculations based on Milanovic (1998). These are not strictly comparable to World Bank PPPs because they are based on an earlier set of price surveys. For some other countries without PPP exchange rates in the World Bank's World Development Indicators online database, we derived PPPs implicitly from World Bank Povcalnet data, http://iresearch.worldbank.org/PovcalNet/index.htm2.

[24] The World Top Incomes Database, constructed by Facundo Alvaredo, Tony Atkinson, Thomas Piketty, and Emmanuel Saez. Data downloaded from http://topincomes.g-mond.parisschoolofeconomics.eu/.

Table 11.2 Coverage of countries and populations with *both* household surveys and PPP data, 1988–2005

Year	Number of countries	Population in billions (% of world population)
1988	92	4.45 (87)
1993	104	5.06 (91)
1998	109	5.32 (90)
2002	113	5.78 (92)
2005	119	5.95 (92)
Total	537	

Source: Authors' calculations.

individuals had an income no higher than what would be expected of a midlevel manager in an international firm. This suggests that very rich households are simply excluded from surveys, which is the assumption we make in incorporating top income data into our survey distributions. In other words, we assume that the survey data in the Milanovic data set represent only the bottom 99% of the population in each country. Accordingly we multiply the population in each income group in the surveys by 0.99 and append the top percentile with its income share from the tax data (assuming that its share of "control" income is equal to its share of survey income). The exclusion of the top percentile implies that mean income in the surveys is underestimated, and our procedure results in a corresponding increase in mean income for each country.

For those country-years that do not have top income data, we impute top percentile shares on the basis of regression. The income share of the top decile in Milanovic's household survey data is strongly correlated with the income share of the top percentile in the independently estimated top income data. Excluding one visible outlier in the 104 country-years with both Milanovic data and top income data,[25] the simple OLS regression coefficient of the income share of the top percentile against the income share of the top decile (on the remaining 103 datapoints) has a *t*-statistic of 7.46 and an R^2 of 0.36. We then added the original mean income from the surveys as a further regressor. Mean income is found to be highly significant with a *t*-statistic of 6.69, the top decile share becomes still more significant with a *t*-statistic of 10.33, and the regression R^2 rises to 0.55.[26] We use this latter regression to generate predicted values for the income share of the top percentile for country-years without tax data.

[25] The outlier is South Africa for 1993, for which the top decile share in Milanovic's data is exceptionally large at 46%, whereas the top percentile share from income tax data, at 10.3%, is much smaller than would be expected given this.

[26] The estimated regression equation is topone $= -6.8 + 0.51$topten $+ 0.30$meaninc where topone and topten are, respectively, the shares of the top percentile and the top decile in percentage points, and meaninc is mean survey income in PPP$ thousand. Year dummies and demographic variables were insignificant.

Lakner and Milanovic (2013) take a different approach to imputing top income shares in estimating global inequality between 1988 and 2008.[27] Following Banerjee and Piketty's (2010) finding in India that a significant part of the discrepancy between estimates of consumption expenditure in the national accounts and in household surveys can be accounted for by missing or underreported top incomes, Lakner and Milanovic attributed the difference between HFCE and survey incomes (when the former is larger than the latter) entirely to the top decile of the national distribution in each country-year, and add this residual to the income of the top decile reported in the survey. They then calculated a Pareto coefficient for each country-year distribution on the basis of the unadjusted survey income in the ninth decile and the adjusted income in the top decile (following the procedure described in Atkinson, 2007). Assuming this Pareto distribution applies *within* the top decile of each country-year distribution, they estimated income shares for the income groups P90–P95 (i.e., percentile 90 to percentile 95), P95–P99, and P99–P100, yielding 12 income groups per country-year.

An implicit assumption behind Lakner and Milanovic's procedure for imputing top incomes is that HFCE per capita is the *correct* measure of mean consumption expenditure (or income), when it is larger than the corresponding survey mean. We have argued against using national accounts means in Section 11.4 and in Anand and Segal (2008). It should also be noted that Milanovic's (2002, 2005, 2012) own previous estimates of "true" global inequality are based on his assumption that survey means are preferable to national accounts means.

11.5.1 Global Inequality Estimates With and Without Top Income Data

Our results for global inequality are presented in Table 11.3 and Figure 11.1. The first notable finding is the very high level of global inequality. Considering the global distribution with top incomes over the period 1988–2005, the Gini varies between 0.722 and 0.735, MLD (or Theil L) between 1.093 and 1.156, and Theil T between 1.114 and 1.206. The top percentile in the world has a share between 17.3% and 20.7% of global income, and the top decile between 58.5% and 62.0%. The richest percentile in the world have mean incomes almost 21 times the world mean income in 2005, or a mean per capita household income of about PPP$90,000 in 2005. The threshold for being in the top percentile in 2005 was PPP$42,000.[28]

As anticipated, the inclusion of top income data raises the estimated levels of inequality relative to those based on household surveys without top income data. The average

[27] Unfortunately, the Lakner and Milanovic data for the benchmark year 2008 were not made available to us for our own calculations of global inequality.

[28] For comparison, the threshold defining the top percentile in the United States in 2005 was a total household income of PPP$342,000, which for a four-person household implies a per capita figure of PPP$85,500, or approximately double the global threshold.

Table 11.3 Global inequality with and without top incomes, 1988–2005

Year	Income share of top percentile (%)	Income share of top decile (%)	Gini	Between-country Gini (% of global Gini)	MLD	Between-country MLD (% of total)	Within-country MLD (% of total)	Theil T	Between-country Theil T (% of total)	Within-country Theil T (% of total)
With top incomes										
1988	17.3	58.5	0.726	0.649 (89)	1.136	0.886 (78)	0.250 (22)	1.114	0.780 (70)	0.334 (30)
1993	17.6	58.5	0.727	0.636 (88)	1.142	0.836 (73)	0.306 (27)	1.115	0.753 (68)	0.362 (32)
1998	19.0	59.5	0.722	0.632 (88)	1.093	0.780 (71)	0.314 (29)	1.145	0.750 (66)	0.395 (34)
2002	20.6	62.0	0.735	0.649 (88)	1.133	0.830 (73)	0.303 (27)	1.206	0.809 (67)	0.397 (33)
2005	20.7	60.0	0.727	0.633 (87)	1.156	0.806 (70)	0.349 (30)	1.188	0.755 (64)	0.433 (36)
Without top incomes										
1988	11.2	54.8	0.705	0.642 (91)	1.063	0.861 (81)	0.202 (19)	0.967	0.764 (79)	0.202 (21)
1993	11.6	54.9	0.707	0.632 (89)	1.069	0.819 (77)	0.250 (23)	0.976	0.745 (76)	0.231 (24)
1998	13.1	56.9	0.698	0.624 (89)	1.008	0.757 (75)	0.251 (25)	0.969	0.732 (76)	0.236 (24)
2002	14.1	58.5	0.711	0.640 (90)	1.046	0.801 (77)	0.245 (23)	1.027	0.788 (77)	0.239 (23)
2005	14.9	56.5	0.701	0.622 (89)	1.060	0.775 (73)	0.285 (27)	0.977	0.725 (74)	0.252 (26)

Source: Authors' calculations.

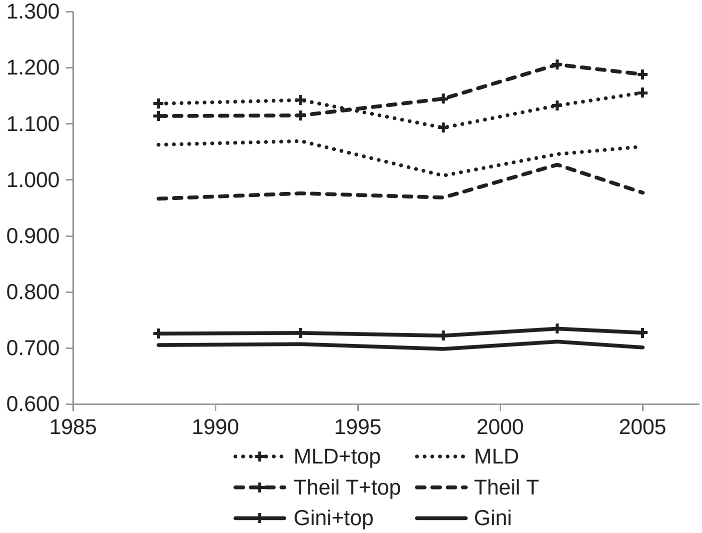


Figure 11.1 Global inequality with and without top incomes, 1988–2005. *Note: Estimates with top income data included are denoted "+top". Source: Table 11.3.*

share of the top percentile over the period 1988–2005 increases from 13.0% on the basis of the surveys alone to 19.0% when top income data are included. Correspondingly, the average top decile share over the period is 56.3% with survey data alone, and 59.7% when top income data are added. Depending on the year, the Gini increases by 3–4%, MLD (or Theil L) by a larger margin of 7–9%, and Theil T by the largest margin of 14–22%. For all measures the increase is greatest in 2005, when the inclusion of top income data raises the global Gini by 4%, MLD by 9%, and Theil T by 22%. These differences in impact reflect the different sensitivities of the measures to income changes at the top end of the distribution.

Turning to changes in inequality with top income data during 1988–2005, the income share of the top percentile rises monotonically from 17.3% to 20.7%. The share of the top decile rises from 58.5% in 1988 to 60.0% in 2005, peaking at 62% in 2002. The Gini coefficient shows very little movement in this period: the highest Gini value is 0.735 (in 2002), which is only 0.013 higher than the lowest Gini value of 0.722 (in 1998), a difference of under 2%. MLD (or Theil L) and Theil T show somewhat larger movements, with the difference between the highest and lowest years for the measures being 6% and 8%, respectively. MLD peaks in 2005 and Theil T in 2002, and for both of these measures inequality rises over the period 1988–2005—for MLD by 1.8% and for Theil T by 6.6%.

The top income data modify both the estimated level of global inequality and its rate of change over time. Although inequality rises over 1988–2005 according to MLD and Theil T with top income data included, inequality is virtually unchanged over the period according to all three measures when top income data are not included. In the latter case the Gini is 0.705 in 1988 and 0.710 in 2005, MLD is virtually unchanged at 1.06, and Theil T rises marginally from 0.967 to 0.977; however, the income share of the top percentile rises from 11.2% to 14.9%.

The changes in inequality over time are not large compared to changes witnessed in some individual countries. This is particularly so in the case of the Gini coefficient, where the peak-to-trough difference is only 1.3 Gini points with the top income data, compared to a rise, for example, of about 5 Gini points in the United States over the period 1988–2005.[29] Moreover, given the different sources of estimation error that we described in Section 11.4, the small changes we find in the global inequality indices may not be statistically significant—particularly in the cases of the Gini and MLD, which are less than 2 percentage points different in 2005 from 1998. However, the rise in the share of the global top percentile, from 17.3% to 20.7% during 1988–2005, seems less trivial; it implies that the incomes of the top percentile increased by 20% relative to mean income—though we note that this is also smaller than the rise in the share of the top percentile in the United States over the same period, from 15.5% to 21.9%.[30]

In the Appendix we provide analogous results for the "common sample over time" of 67 countries. Whereas the full sample shown earlier comprises between 87% and 92% of the world population depending on benchmark year, the common sample over time comprises between 79% and 82%. As can be seen in Appendix Table 11.A2, the global inequality estimates are very similar to those in Table 11.3 shown earlier. The Gini coefficient for the common sample is never more than 1 percentage point different from that for the full sample, whereas MLD and Theil T are never more than 3 percentage points different. Note that the common sample over time is not necessarily more representative of the global income distribution than our full sample in each benchmark year, and estimates of the level or rate of change of global inequality based on the common sample are not necessarily more accurate.

Our calculations with top income data assume that household surveys do not capture the top percentile of the national income distribution. An alternative way to include the top income data is to assume that surveys are indeed representative of all households, but that they underreport the incomes of the top percentile in the national distribution. This is the assumption made by Alvaredo and Londoño Vélez (2013) and requires a different calculation. Rather than multiply the population of each income group in the survey data by 0.99 and then append the top percentile with its income share from the tax data, on the alternative assumption one simply *replaces* the income of the top percentile in the survey data with that from the tax data. We have performed this calculation as well, and it leads

[29] See Atkinson et al. (2011, p. 33, Figure 6), series adjusted with tax data including capital gains.

[30] This refers to the share of the top percentile including capital gains, downloaded from the World Top Incomes Database, http://topincomes.g-mond.parisschoolofeconomics.eu/.

to marginally lower estimates of global inequality: in the five benchmark years the global Gini is upto 0.4% smaller, and MLD and Theil T are upto 1.2% smaller. However, for the latter two decomposable indices, the within-country component is noticeably smaller, by 3.6–5.2% for MLD and by 2.4–4.1% for Theil T—but the between-country component is about 0.5% larger for both indices.

11.5.2 Comparison of Alternative Estimates of Global Inequality

We saw earlier that only three previous studies have used 2005 PPPs to estimate global inequality: Bourguignon (2011), Milanovic (2012), and Lakner and Milanovic (2013). Milanovic's (2012) estimates of global inequality are directly comparable to our estimates without top incomes in Table 11.3, as they are based on the same survey data and methodology. The only substantial difference we know of is in the PPPs used for countries for which the World Bank does not have data (see footnote 23 for the sources that we use in these cases). Milanovic found the Gini coefficient to vary between 0.684 and 0.707 in the period 1988–2005, whereas in our estimates given earlier it varies between 0.698 and 0.711. However, whereas we find Theil T at virtually the same level in 1988 as in 2005, he found it to rise from 0.875 to 0.982 over the same period.

Lakner and Milanovic (2013), like us, estimated global inequality both with and without imputed top incomes. Their estimates without top incomes also follow the same methodology as Milanovic (2012) and are based directly on survey data. Lakner and Milanovic's estimates of the global Gini without top incomes are close to ours, varying between 0.705 and 0.722 in the period 1988–2008. Their Theil T is slightly higher than ours, varying between 1.003 and 1.049 in the period. Significantly, their MLD shows a marked decline, from 1.142 in 1988 to 1.027 in 2008.

Lakner and Milanovic—like us—found that imputing top incomes leads to higher estimates of global inequality. Their HFCE-based method of imputing top incomes, discussed earlier, raises the global Gini by 3.8–6.3 Gini points, with the difference rising over time in the period 1988–2008.[31] Nonetheless, their Gini ends the period at almost exactly the same level as it began, declining marginally from 0.763 in 1988 to 0.759 in 2008. This is a much larger effect than we find from adding top income data to the survey data. As we saw in Table 11.3, our method leads to the Gini being approximately 2 Gini points higher in each year. Lakner and Milanovic themselves pointed out that their imputation assumption is rather extreme in some cases. For example, in 2008 in India—the country that appears to have motivated their procedure—they find the survey mean to be only 53% of HFCE per capita, so they attribute the remaining 47% of total HFCE *entirely* to the top decile, adding it to the income of the top decile reported in the survey. This adjustment seems implausibly large to us. Conversely, for China in both 1988 and 2008,

[31] Lakner and Milanovic do not give estimates of other inequality measures for their distribution with imputed top incomes.

HFCE is smaller than survey income, so no adjustment is made by the authors for under-reporting or undersampling of top incomes.

The final study that uses 2005 PPPs to estimate global inequality is Bourguignon (2011), which—unlike the other studies mentioned in this section—scales within-country distributions to GDP per capita. Bourguignon found the Gini coefficient to decline from approximately 0.70 to 0.66 between 1989 and 2006 (these numbers were read off his Figure 1). This is a substantial decline compared with the findings reviewed earlier of virtually no change in the Gini without top incomes. The main difference between Bourguignon's estimates and the other estimates without top incomes discussed here is that Bourguignon scales to national accounts data. In Section 11.4 we argued that if one uses national accounts data then HFCE is preferable to GDP as an approximation to household income, so we compare estimates based on survey means with estimates based on HFCE means in the next section.

11.5.3 Global Inequality Estimates Using NA Means, Without Top Income Data

In this section we report global inequality estimated by scaling household survey incomes so that the scaled mean is equated to per capita HFCE from NA in each country (in contrast to using incomes directly from the surveys). HFCE figures in PPP$ are not available for all the country-years for which we have household survey data. In each year, we distinguish between the "full sample" defined as the set of all countries with survey data, and the "common sample" *across data sets* defined as the subset of the full sample countries that *also* have HFCE data in PPP$ (note that this is *different* from the "common sample over time," defined earlier). In 1988 the common and full samples are quite different: in that year the countries that have *both* survey data and HFCE data in PPP$ comprise only 77% of the world population, compared with 87% of the world population for countries in the full sample (see Table 11.2).[32] In the other years covered in Table 11.2, the common sample has 3–4% less of the world population than the full sample.[33]

For each of our indices, Table 11.4 presents three different global inequality estimates without top income data: first, the full sample estimates as in Table 11.3; second, estimates based on survey data restricted to the common sample; and third, the common sample estimates based on per capita HFCE (as described earlier). We will refer to the first as the

[32] For some countries where World Development Indicators (WDI) does not have HFCE data in PPP$, HFCE is nevertheless available in local currency units (LCUs). By definition, these countries do not have PPP exchange rates in WDI, but for 11 of them we have PPPs (see footnote 23), which we use with the survey data. However, in almost all cases using these PPP conversion rates gives implausible results for HFCE. For this reason we do not use these data.

[33] For the years 1993, 1998, 2002, and 2005 the common sample covers 87%, 87%, 89%, and 89%, respectively, of the world population compared to the full sample percentages in Table 11.2.

Table 11.4 Global inequality estimates using survey means and HFCE means, without top incomes, 1988–2005

	Gini			MLD			Theil T		
Year	Survey means, full sample	Survey means, common sample	HFCE means, common sample	Survey means, full sample	Survey means, common sample	HFCE means, common sample	Survey means, full sample	Survey means, common sample	HFCE means, common sample
1988	0.705	0.721	0.739	1.063	1.148	1.243	0.967	1.024	1.094
1993	0.707	0.706	0.721	1.069	1.074	1.133	0.976	0.970	1.036
1998	0.698	0.699	0.711	1.008	1.010	1.062	0.969	0.973	1.017
2002	0.711	0.710	0.706	1.046	1.039	1.032	1.027	1.022	1.016
2005	0.701	0.698	0.698	1.060	1.044	1.024	0.977	0.967	0.984

Note: Full sample consists of all countries for which household survey data are available, as in Tables 11.2 and 11.3. Common sample in a given year consists of the subset of countries with *both* survey data and HFCE data in PPP$ in that year.
Source: Authors' calculations.

full sample survey-means estimates, the second as the common sample survey-means estimates, and the third as the common sample HFCE-means estimates. This highlights the fact that in the first two, mean incomes in each country are obtained directly from the surveys, whereas in the third the mean is externally imposed from HFCE data. These estimates exclude the top income data so that we can focus on the differences in global inequality using survey and NA mean incomes. Figure 11.2 plots the Gini coefficients for the three different sets of data.

The most notable difference between the full sample survey-means estimates and the common sample HFCE-means estimates is that whereas the former appear relatively flat, the latter have a clear downward trend. The two estimates are approximately the same in 2002 and 2005, but because of their different starting points in 1988 the full sample survey-means Gini declines by only 0.004 between 1988 and 2005, from 0.705 to 0.701, whereas the common sample HFCE-means Gini declines by 0.041, from 0.739 to 0.698. However, the common sample survey-means estimates indicate that about half this difference is explained by the difference between the full and common sample: the common sample survey-means Gini declines by 0.023, from 0.721 to 0.698.

The second factor that appears to explain the difference in trend for the HFCE-means Gini and the survey-means Gini is the divergent trend for India specifically in comparing survey and HFCE means, a phenomenon that has been examined in detail by Deaton and Kozel (2005). In our data the average annual growth rate of per capita household consumption expenditure in India from 1988 to 2005 was 2.8% according to surveys, and more than twice that at 5.8% according to HFCE from the National Accounts. When we both restrict the estimates to the common sample *and* exclude India from it, the survey-means Gini declines by 0.029, whereas the HFCE-means Gini decreases by a

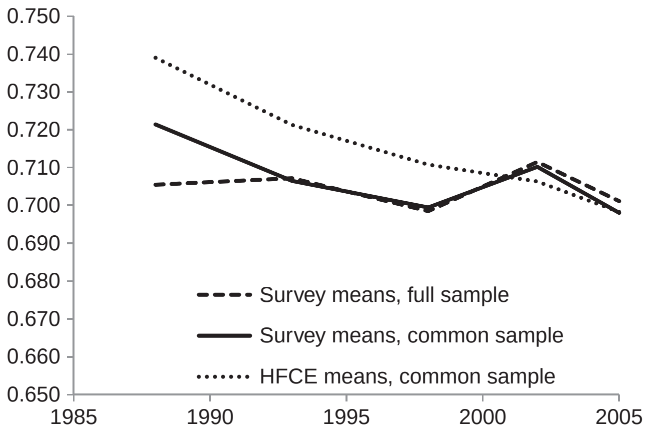


Figure 11.2 Global Gini without top incomes, using survey means and HFCE means, 1988–2005. *Note: Common sample consists of surveys restricted to country-years with HFCE data. Source: Table 11.4.*

similar magnitude of 0.034.[34] Thus ensuring a common sample and excluding India virtually eliminates the difference in trend between HFCE-means and survey-means estimates of the Gini.[35]

Thus the divergence between HFCE-means and survey-means estimates seems to be due to the loss of as much as 10% of the world population in the "common sample" that has *both* survey and HFCE data, and the divergent trends in India. Given this, in our view the decline in global inequality implied by the HFCE calculations is likely to be illusory. We have not examined estimates based on GDP means, as opposed to HFCE means, for the reasons mentioned earlier. Still, these findings do not seem to corroborate Bourguignon's (2011) result for the period 1989–2006, based on GDP data and discussed earlier, that "inequality decreases, and it decreases at a very fast pace."

11.6. BETWEEN- AND WITHIN-COUNTRY INEQUALITY

Table 11.3 also presents estimates of between-country and within-country inequality, and Figure 11.3 plots these estimates. Between-country inequality is defined as global inequality under the hypothetical assumption that every individual is assigned his or her country's mean per capita household income. It suppresses inequality within countries and measures inequality in the global distribution among world citizens where the only source of variation is mean per capita income across countries (in other words, between-country inequality is just concept 2 global inequality). Between-country inequality is well-defined for any inequality index, and we report it in Table 11.3 for the Gini, MLD (or Theil L), and Theil T measures. For the decomposable measures MLD and Theil T, the difference between overall global inequality and between-country inequality is a weighted average of inequality in each country, and is denoted as *within-country inequality*. In the case of MLD (i.e., Theil L), within-country inequality is a population-share weighted average of the MLD in each country, whereas for Theil T it is an income-share weighted average of the Theil T in each country (Anand, 1983, pp. 86–92).

In the case of MLD (Theil L) only, the within-country component has an additional interpretation: it is equal to what global inequality would be under the hypothetical assumption that mean per capita incomes are equalized between countries, while relative inequality is kept constant within each country. In this sense it is a natural complement to the definition of between-country inequality, and for this reason we consider MLD to be strictly decomposable, but Theil T to be only weakly decomposable (Anand, 1983, pp. 198–202).

[34] The levels are different, however, with the survey-means Gini declining from 0.701 to 0.671 and the HFCE-means Gini declining from 0.722 to 0.689.

[35] This procedure also substantially reduces the difference in trend for MLD and Theil T, but does not eliminate it as effectively as with the Gini.

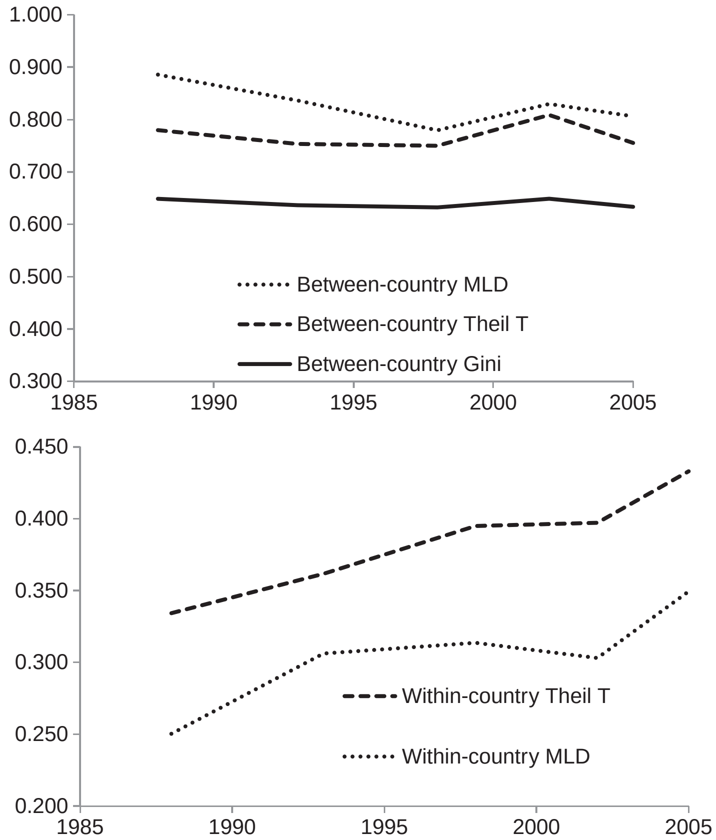


Figure 11.3 Between-country and within-country global inequality with top incomes, 1988–2005. *Source: Table 11.3.*

Considering estimates of global inequality with top incomes, we make four observations from Table 11.3 (top panel). First, between-country inequality is larger than within-country inequality for both the decomposable indices. Between-country inequality ranges between 70% and 78% of overall global inequality for MLD and between 64% and 70% for Theil T.

Second, the inclusion of top income data increases the within-country component substantially, as would be expected. For MLD the within-country component rises by between 23% and 25%, depending on the year, whereas for Theil T it rises by between 57% and 72%. The between-country component also changes because our imputation of the income share of the top percentile increases country mean incomes by different proportionate amounts.

Third, from 1988 to 2005 between-country inequality declines by all three measures, as shown in Table 11.3. For the estimates with top income data, the between-country

Gini falls by 2% from 0.649 to 0.633, the between-country MLD declines 9% from 0.886 to 0.806, whereas the between-country Theil T declines 3% from 0.780 to 0.755.

Fourth, over the period 1988–2005 within-country inequality clearly increases for both decomposable indices as seen in Figure 11.3. For estimates with top income data, the within-country MLD rises by 40% from 0.250 to 0.349, and the within-country Theil T rises by 30% from 0.334 to 0.433.[36]

The Gini coefficient is not a decomposable measure in either the weak or strong sense. Although we can define the between-country Gini straightforwardly, the residual from overall global inequality cannot be interpreted as within-country inequality (see Anand, 1983, pp. 311–326). However, as with any inequality index, we can answer the question of what happens to the global Gini and to Theil T when country mean incomes are equalized but relative inequality is kept constant within each country (Anand, 1983, p. 201). This question is relevant in assessing the following claims.

On the basis of the fact that between-country inequality is greater than within-country inequality, Sala-i-Martín (2002, p. 39) stated that "the best strategy to reduce world income inequalities is to induce aggregate economic growth in poor countries." Similarly, Rodrik (2013, p. 12) noted that "the more rapid growth of poor countries since the 1990s is the key behind the recent decline in global inequality," concluding from this that "aggregate economic growth in the poorest countries is the most powerful vehicle for reducing global inequality." For economic growth in poor countries to reduce global inequality, it would of course have to be more rapid than growth in richer countries. In this case, the greatest reduction in global inequality that could possibly be achieved without addressing within-country inequality is calculated by eliminating between-country income differences while keeping inequality within each country constant. Conducting this exercise for 2005 with top income data, the Gini would decline from 0.727 to 0.437 and Theil T from 1.188 to 0.433; in the case of the strictly decomposable MLD, the decline is from 1.156 to its within-country component of 0.349. This is a large decline, but global inequality would still remain at about the level of a high-inequality country such as China, where in 2005 we find the Gini to be 0.430, MLD to be 0.367, and Theil T to be 0.324.

In Section 11.3 we pointed out that the concept of sigma convergence in the growth literature has little relationship to any other concept of global inequality. Bourguignon et al. (2004) used GNI per capita and found that what we call concept 2 inequality fell between 1980 and 2002, while concept 1 inequality rose. Similarly, in our data, between-country inequality (i.e., concept 2 global inequality) declines by all measures during 1988–2005, whereas we find "sigma divergence" when we calculate concept 1 global

[36] It should be noted that changes over time in population shares of countries in the case of MLD, and income shares of countries in the case of Theil T, will lead to changes in "within-country inequality" even holding inequality constant *within* each country.

Table 11.5 Concept 1 inequality, calculated using per capita incomes from survey data with top incomes

Year	Gini	MLD	Theil T	Sigma (std. dev. of log-income)	World mean per capita income (PPP$)	China mean per capita income (PPP$)
1988	0.501	0.538	0.414	1.15	3424	342
1993	0.535	0.574	0.480	1.12	3683	526
1998	0.552	0.594	0.523	1.11	3923	863
2002	0.575	0.655	0.573	1.15	4148	1042
2005	0.578	0.665	0.580	1.17	4364	1916

Source: Authors' calculations.

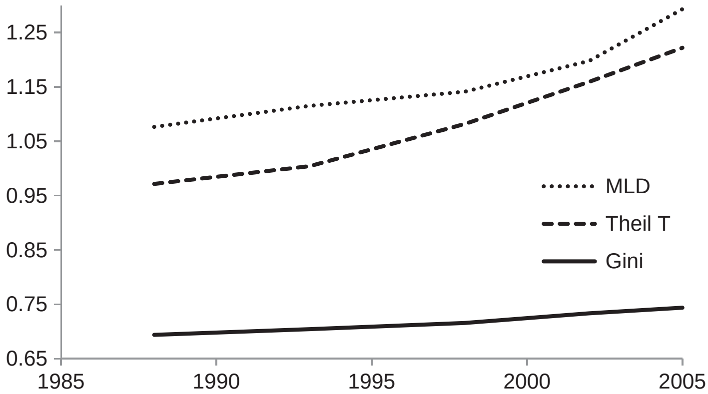


Figure 11.4 Global inequality without China, based on survey data with top incomes. *Source: Authors' calculations.*

inequality, as shown in Table 11.5. The three inequality measures increase when applied to the concept 1 distribution: the concept 1 Gini increases from 0.501 to 0.578, MLD from 0.538 to 0.665, and Theil T from 0.414 to 0.580. The standard deviation of (unweighted) log mean income also rises from 1.15 to 1.17.[37]

Table 11.5 also presents mean per capita survey incomes for the world, and for China separately. Several papers have estimated global inequality excluding China (e.g., Milanovic, 2012; Sala-i-Martín, 2006; and Schultz, 1998), and we present our estimates in Figure 11.4, which include the top income data. They indicate that global inequality without China increases by all three measures: the Gini rises by 0.050, MLD by 0.217 and Theil T by 0.250. We would note, however, that although these estimates are

[37] These figures use per capita incomes calculated from survey data with top incomes. Without the top income data, the surveys alone imply much the same trend for concept 1 inequality: the Gini rises from 0.503 to 0.576, MLD from 0.543 to 0.663, Theil T from 0.417 to 0.576, and sigma again from 1.15 to 1.17.

instructive from the point of view of accounting for global inequality and its evolution, they have no global welfare implications because they exclude approximately one-fifth of the world population.

11.7. RELATIVE AND ABSOLUTE GLOBAL INEQUALITY

In an article entitled "What Are We Trying to Measure?" the development economist Dudley Seers (1972, pp. 34–35, endnote 2) made a telling distinction between relative and absolute inequality when he wrote: "Suppose, for example, that a perspective plan specified that [the] per capita income of Brazil doubled in the next thirty years, but assumed no change in distribution or in the proportion unemployed. Then at the turn of the century, a big landowner in the Matto Grosso could run four cars, instead of two, and a peasant in the North-East could eat two kilogrammes of meat a year instead of one. His son might well be still out of work. Could we really call that 'development'?". Although relative inequalities in this example have remained unchanged, the absolute differences have grown in proportion to the expansion of the economy. In addition to considering relative global inequality, there is clearly a case for examining absolute global inequality—as noted, for example, by Ravallion (2004).

The first thorough investigation of measures of absolute global inequality is by Atkinson and Brandolini (2010). They posited a "world social welfare function," which exhibits a changing social marginal valuation of income at different points along the global income distribution (see also Anand and Sen, 2000). The absolute cost of inequality is then expressed in terms of Atkinson's (1970) concept of "equally distributed equivalent income" for this social welfare function (see also Kolm, 1969). For any income distribution, Atkinson defines the equally distributed equivalent income as that level of income per head, which, if equally distributed, would yield the same level of social welfare as the existing distribution. Then the absolute cost of inequality is the income per head that is "wasted" as a result of inequality (i.e., it is mean income minus the equally distributed equivalent income). (The relative cost of inequality is the absolute cost divided by the mean, which is the definition of Atkinson's index of relative income inequality.)

For the Gini welfare function, the absolute cost of inequality is mean income μ multiplied by the Gini coefficient G, and the relative cost is simply G (Anand, 1983; Sen, 1973). In general, mean income μ times a relative inequality measure produces the corresponding absolute inequality measure. For the relative global inequality measures G, MLD, and Theil T, we also estimate the absolute global inequality measures μG, μMLD, and μT, respectively, where μ is the world mean income. The world mean income at 2005 PPP$ is shown in Table 11.6 for the years 1988–2005.

Table 11.7 shows the evolution of absolute global inequality with top income data between 1988 and 2005 as measured by the Gini, MLD, and Theil T, expressed as a ratio of 2005 world mean income calculated from surveys with top incomes added (PPP$4364).

Table 11.6 World mean income at 2005 PPP$, alternative estimates

Year	World GNI per capita	World mean income from surveys, with top incomes	World mean income from surveys, without top incomes
1988	6433	3424	3115
1993	6524	3683	3342
1998	7262	3923	3514
2002	7934	4148	3713
2005	8772	4364	3865

Source: GNI per capita from *World Development Indicators*, http://data.worldbank.org/data-catalog/world-development-indicators. Estimates of world mean income from surveys, with and without top incomes, are authors' calculations.

Table 11.7 Absolute global inequality using survey data with top incomes added, as a *ratio* of 2005 world mean income (PPP$4364)

Year	Absolute Gini	Absolute MLD	Absolute Theil T
1988	0.569	0.891	0.874
1993	0.614	0.964	0.941
1998	0.649	0.983	1.029
2002	0.698	1.077	1.146
2005	0.727	1.156	1.188

Source: Authors' calculations.

Over the 17-year period 1988–2005, there has been an unambiguous rise in absolute global inequality according to all three measures. This is unsurprising given the rise in world mean incomes over this period. To prevent a rise in absolute inequality, relative inequality has to decrease at a faster rate than the rise in mean incomes—which seems an unlikely prospect for the global economy.

In Section 11.2 of this chapter we noted and discussed the widespread concern about global income inequality—in terms of both its level and change. Given that there appears to be little movement in relative global inequality (see Section 11.5), whereas there is a significant widening of absolute global inequality (Table 11.7), the widespread concern about inequality may be based on people making comparisons of living standards in *absolute* rather than relative terms.

11.8. GLOBAL POVERTY

11.8.1 Methodology

Like absolute global inequality, global poverty is a measure based on absolute living standards. To measure global poverty, an absolute poverty line is applied to the global distribution of income and the number of individuals below it calculated. This procedure is employed to monitor global poverty over time—including for the first Millennium Development Goal. Chapter 9 of this volume discusses poverty in developing countries and regions using different poverty lines. The most widely quoted estimates of global poverty for an absolute poverty line are those produced by Chen and Ravallion (2008, 2012) at the World Bank. The Millennium Development Goal refers to consumption poverty, but the limited availability of surveys around the world and over time necessitates that Chen and Ravallion use a mixture of (consumption) expenditure and income surveys.

Chen and Ravallion's (2008, 2012) estimates for the World Bank use the 2005 ICP PPPs, which, as discussed earlier, are preferable to PPPs based on the earlier ICP rounds. Theirs are the only estimates of global poverty based on unit record data from surveys, which are not publicly released, and which are clearly preferable to the grouped data that are available to other researchers. The World Bank's methodology has been criticized for not scaling the survey data to national accounts (NA) means.[38]

In Section 11.4 on data we discussed the question of whether to use NA means or survey means in the context of estimating global inequality. The methodology used by the World Bank to measure global poverty, like that used by us to estimate global inequality, uses survey data directly to estimate income (or consumption) levels—converted into international dollars using consumption PPPs. As in the case of global inequality, some authors have calculated global poverty by using survey data for within-country relative distributions and NA data for country mean incomes.[39] In the context of global inequality we argued that using means directly from surveys is preferable to scaling them to NA levels, and those arguments apply even more for measuring global poverty.

There is a further consideration that makes scaling to NA categories even less appropriate for estimating global poverty. We know that surveys tend to exclude very rich households and/or underreport their incomes, and for this reason they are likely to underestimate mean income or consumption. But this implies that scaling up the income (consumption) of *every* household to ensure that the survey mean is made equal to the NA mean will imply overestimating the income (consumption) of *all* but the richest households. Put another way, the "missing" income of the rich will be inappropriately divided among the entire population. Poverty will therefore be underestimated (for further discussion, see Anand et al., 2010, pp. 13–14).

Turning to the choice of poverty line, the World Bank uses what is commonly known as the "$1-a-day" line. This was originally defined in World Bank (1990) as PPP$1-per-day at 1985 PPP. This poverty line was chosen informally as being representative of the poverty lines of the poorest countries, converted into 1985 PPP$.

The difficulty arises over how to update a 1985-based PPP$ value.[40] Within a single country, one would usually update a poverty line by using a price index based on

[38] For example, see Sala-i-Martín (2006).
[39] See Sala-i-Martín (2006).
[40] This discussion draws on Anand et al. (2010).

measured inflation. Because the international PPP dollar is indexed to the US dollar in the ICP benchmark year, one might think that all we need to do is to deflate by the US inflation rate. But updating a poverty line that is denominated in PPP$ is not so simple. As discussed earlier, calculating a set of PPP exchange rates involves the prices of all countries, so that changes in a country's PPP exchange rate will depend on price changes in all countries. Bangladesh's 1985 poverty line in 1985 PPP$, adjusted for US inflation during 1985–1993, would not be expected to be equal to Bangladesh's 1993 poverty line in 1993 PPP$.[41]

The World Bank updated the 1985 global poverty line to PPP$1.08 a day at 1993 PPP and now uses PPP$1.25-a-day at 2005 PPP. This represents a lower rate of inflation than in the US, but as Chen and Ravallion (2001, p. 288) pointed out: "the fact that $1.08 in 1993 has a US purchasing power less than $1 in 1985 does not mean that the real value of the poverty line has fallen. Indeed, if we had simply adjusted the $1 per day line for inflation in the US between 1985 and 1993 we would have obtained a poverty line which is well above the median of the ten lowest poverty lines at 1993 PPP." The World Bank has chosen consistency with those domestic poverty lines as the most important criterion in setting a global poverty line. This can be justified by arguing that domestic poverty lines will have maintained their real value within their respective countries better than a PPP inflation-adjusted measure. Therefore each time they updated the poverty line, it was derived as the median of the lowest 10 poverty lines in their data set converted into PPP$ from the most recent ICP (Chen and Ravallion, 2001, 2008). Although this has an obvious logic to it, Deaton (2010) pointed out that the fact that the composition of the bottom 10 countries will change over time can lead to inconsistency: India exited the bottom 10 countries in their data set in the 2005 update owing to its relatively high growth rate, and because its poverty line was relatively low for its income level, this exit led to a rise in the poverty line relative to where it would be with India in the bottom 10 countries. This has the paradoxical implication that a rise in India's income can lead to a rise in estimated global poverty.

A more fundamental challenge to the PPP$1-a-day poverty line has been made by Reddy and Pogge (2010). They objected to the money metric approach to global poverty measurement, noting that the PPP$1-a-day poverty line does not correspond to any "achievement concept" or set of capabilities that are common across countries. That is, there is no reason to think that PPP$1-a-day in one country will enable the same set of achievements—for example, in terms of nutrition or shelter—as PPP$1-a-day in another country. Although domestic poverty lines are often set according to some achievement concept, this interpretation is lost when a global poverty line is constructed

[41] More generally, national income at PPP$ calculated in year $t+n$ is not equal to national income at PPP$ calculated in year t multiplied by intervening domestic growth and deflated by intervening U.S. inflation (see Anand et al., 2010, p. 6).

using standard PPP exchange rates. Reddy and Pogge argued that an explicit achievement-based threshold should be used to define a global poverty line. This would require costing a minimal basic set of capabilities in each country to yield a money-metric poverty line denominated in local currency. Thus the global capability-based poverty threshold would be represented in income space by the set of these national poverty lines, one for each country.[42] Although this has theoretical attractions, it has not been implemented in practice.

11.8.2 Poverty Estimates

The question of updating the global poverty line remains contentious, and the World Bank reports poverty headcounts for several different poverty lines. Its poverty measurement website Povcalnet[43] also allows the user to choose a poverty line for which it then provides estimates. The latest official publication of global poverty numbers (Chen and Ravallion, 2012) presents poverty headcounts using the following poverty lines all at 2005 PPP$: PPP$1-a-day, which they describe as "close to India's (old) national poverty line" and "an exceptionally frugal line even by the standards of the world's poorest countries" (Chen and Ravallion, 2012, p. 1); PPP$1.25-a-day, the line derived from domestic poverty lines in poor countries as described earlier; and PPP$2-a-day. We report the World Bank estimates in Tables 11.8 and 11.9.

Table 11.8 indicates that the first Millennium Development Goal, which was to halve the percentage of people living below "PPP$1-a-day" (i.e., PPP$1.25 at 2005 PPP) from its 1990 level by 2015, was almost achieved by 2008: this percentage declined from 43.1% of the developing world to 22.4%. Chen and Ravallion (2012) reported that the goal was in fact achieved in 2010, though the data for 2010 are not fully representative. If the poverty line of PPP$1-a-day at 2005 prices is used, then the goal was fully achieved in 2008, with the percentage of the poor in developing countries falling from 30.8% in 1990 to 14.0% in 2008 (Table 11.8).

It is clear from Table 11.9 that the distribution of this decline in poverty was highly uneven. The poverty rate in sub-Saharan Africa at the PPP$1.25 line was only slightly lower in 2008 than in 1981, at 47.5% rather than 51.5%. It is little comfort that the 2008 figure for this region is a larger drop from the 1993 peak of 59.4%. As is well known, a large share of the decline in global poverty is due to China, which managed to cut poverty from 84% in 1981 to 60.2% in 1990 and to 13.1% in 2008. The world excluding China

[42] One could then use the relative cost of this capability set in different countries to infer the implied "PPP" exchange rates. There is no reason to think that such exchange rates would be similar to extant PPP exchange rates.

[43] http://iresearch.worldbank.org/PovcalNet/index.htm?2.

Table 11.8 World bank global poverty estimates, 1981–2008

	PPP$1		PPP$1.25			PPP$2	
Poverty line	Number (millions)	Percent of developing world	Number (millions)	Percent of developing world	Number excl China (millions)	Number (millions)	Percent of developing world
1981	1545.3	41.6	1937.8	52.2	1102.8	2585.3	69.6
1984	1369.3	34.7	1857.7	47.1	1137.8	2680.0	68.0
1987	1258.9	30.1	1768.2	42.3	1182.5	2710.2	64.8
1990	1364.7	30.8	1908.6	43.1	1225.5	2864.1	64.6
1993	1338.1	28.7	1910.3	40.9	1277.6	2941.5	63.1
1996	1150.0	23.5	1704.0	34.8	1261.2	2864.8	58.6
1999	1181.9	23.1	1743.4	34.1	1297.0	2937.9	57.4
2002	1096.5	20.6	1639.3	30.8	1276.2	2848.4	53.5
2005	886.1	16.0	1389.6	25.1	1177.7	2595.8	46.9
2008	805.9	14.0	1289.0	22.4	1116.0	2471.4	43.0

Note: All three poverty lines of PPP$1, PPP$1.25, and PPP$2 are at 2005 PPP.
Source: Chen and Ravallion (2012).

Table 11.9 Headcount index of poverty (%) by region, 1981–2008, for poverty line of PPP$1.25 at 2005 PPP

Region	1981	1984	1987	1990	1993	1996	1999	2002	2005	2008
East Asia and the Pacific	77.2	65.0	54.1	56.2	50.7	35.9	35.6	27.6	17.1	14.3
China	84.0	69.4	54.0	60.2	53.7	36.4	35.6	28.4	16.3	13.1
Eastern Europe and Central Asia	1.9	1.6	1.5	1.9	2.9	3.9	3.8	2.3	1.3	0.5
Latin America	11.9	13.6	12.0	12.2	11.4	11.1	11.9	11.9	8.7	6.5
Middle East and North Africa	9.6	8.0	7.1	5.8	4.8	4.8	5.0	4.2	3.5	2.7
South Asia	61.1	57.4	55.3	53.8	51.7	48.6	45.1	44.3	39.4	36.0
Sub-Saharan Africa	51.5	55.2	54.4	56.5	59.4	58.1	58.0	55.7	52.3	47.5
Total developing world	52.2	47.1	42.3	43.1	40.9	34.8	34.1	30.8	25.1	22.4
Total developing world excluding China	40.5	39.1	38.1	37.2	36.6	34.3	33.6	31.5	27.8	25.2

Source: Chen and Ravallion (2012).

also succeeded in reducing poverty, but at a much slower rate. Excluding China, global poverty declined by less than a third since 1990, from 37.2% to 25.2%.

One notable feature of the global distribution of poverty is that much of it is found outside the poorest countries. For instance, in 2008 India was a lower-middle-income country,[44] yet it contained approximately 380 million people below the PPP$1.25 poverty line,[45] or 30% of the global total and about the same number as in all of sub-Saharan Africa.[46] Indeed, Sumner (2012) points out that a majority of people below the World Bank poverty line live in middle-income countries.

In Section 11.6 we noted that within-country inequality has risen over the past two decades, suggesting that the decline in global poverty has been driven by aggregate growth in low- and middle-income countries. It does not follow, however, that continued aggregate growth is the only way to continue to reduce poverty. Redistribution within countries could also play a significant role in poverty reduction in poor countries, just as it does in rich countries.[47]

The average annual income of those living below the PPP$1.25-a-day poverty line was PPP$421 in 2005.[48] Using our estimated global income distribution presented earlier, the individuals in the richest percentile of the world, whose annual incomes averaged PPP$90,000, were therefore 214 times richer than individuals in the poorest 21% of the world. Put another way, the richest 1% of the world, or 65 million people, had a *total income* a little more than 10 times that of the poorest 21%, or 1390 million people.

11.9. CONCLUSION

Rising top income shares in many countries have pushed inequality up the public agenda, while globalization has given this concern a global reach: people no longer compare their lot only to those within their own country. Moreover, the global financial crisis and recession have made the interconnectedness of people's material well-being around the world all the more obvious. Hence the global distribution of income, global inequality, and global poverty are increasingly in the public view.

[44] The World Bank defines lower-middle income countries in terms of US$, not PPP$. They are defined as those countries with per capita GNI between US$1036 and US$4085 in 2012 prices, calculated using the Atlas method. India's per capita GNI in 2008 was US$1050 in current prices, or US$1120 in 2012 prices (inflating by the US CPI). In PPP$, its per capita GDP in 2008 was PPP$2900 in current prices.

[45] This is calculated by applying the poverty headcount for 2009 from Povcalnet (the closest year available) to the 2008 population (http://iresearch.worldbank.org/PovcalNet). Applying the poverty headcount reported in Table 11.9 to sub-Saharan Africa's population in 2008 implies 389 million poor in that region.

[46] For this reason Collier's (2007) book *The Bottom Billion*, which is about a set of poor countries whose populations sum to about one billion, is misleadingly titled: those one billion individuals are not the poorest billion people in the world (Segal, 2008).

[47] In the EU15 countries, for example, 16% of the population were living below their respective national poverty lines in 2003, a figure that would rise to an estimated 25% in the absence of cash benefits that comprise a total of 6.6% of GDP (Guio, 2005). Ravallion (2009) considers the capacity of targeted redistribution within developing countries to eliminate poverty, asking which countries could eliminate poverty by taxing only those with incomes above PPP$13-a-day. Segal (2011) considers the impact on poverty of a universal unconditional transfer, or "basic income," funded by a country's own natural resource rents.

[48] This was calculated from a reported poverty gap index of 7.78% in Povcalnet.

A number of studies have estimated the global distribution of income using a variety of data and methods. Recent advances in data collection have provided us with a much more detailed and accurate view of the global distribution of income than was possible even a decade ago. Household surveys now cover the vast majority of the population of the world, whereas intercountry real income comparisons have been greatly improved with the 2005 round of the ICP. This chapter has also highlighted and used the additional information provided by the growing database of top incomes from tax records. Because survey data typically underestimate or underreport the incomes of the very rich, we have estimated global inequality by appending top income data to the available survey data.

This chapter has described the conceptual foundations of the analysis of the global distribution of income, the confusions that can arise by conflating different concepts of that distribution, and the divergent inequality trends that they can display. Its main focus was the global interpersonal distribution of income, or the concept 3 global distribution. Implicitly this analysis assumes a cosmopolitan symmetric social welfare function, according to which the country or location of an individual in the world is irrelevant.

Our calculations show that when the global interpersonal distribution of income is estimated through household survey data *without* top incomes, inequality in this distribution is very high but remains virtually unchanged from 1988 to 2005. The Gini is 0.705 in 1988 and 0.701 in 2005, MLD decreases slightly from 1.063 to 1.060, and Theil T rises marginally from 0.967 to 0.977. However, the income share of the top percentile rises from 11.2% to 14.9%. The equivalent estimates by Milanovic (2012) and Lakner and Milanovic (2013) also find the Gini virtually unchanged at much the same level as us, but the former finds a rise in Theil T, whereas the latter find a fall in both MLD and Theil T.

We argued that the method used in all of these estimates, which take household incomes directly from surveys, was preferable to the method of "scaling" within-country distributions to NA means. Moreover, we argued that if one were to scale in such a way, then HFCE would be preferable to GDP. Bourguignon (2011) scaled within-country distributions to per capita GDP and found a substantial decline in the global Gini coefficient during 1989–2006, which we also find when we use HFCE. However, we find that the divergence in global inequality estimated using HFCE-means and survey-means is due to a reduced coverage of HFCE data relative to survey data in the first year, 1988, and to the sharp divergence between the survey and HFCE means in India in particular.

When we append data on top incomes to the survey distributions, including imputed incomes for countries without such data, we find that inequality is higher, and there is some indication of an increase in global inequality. In this case the Gini, which is less sensitive to inequality at the top end of the distribution, remains virtually unchanged over the period, at a higher level of 0.722 to 0.735. But MLD and Theil T increase over the period—MLD only slightly from 1.136 to 1.156, and Theil T from 1.114 to 1.188. Given the diverse sources of potential error in estimates of the global distribution of income, these changes may not be statistically significant. A larger proportional rise occurs in the income share of the top percentile—from 17.3% to 20.7%. Thus the increase in estimated global inequality over time with top income data added appears to be driven by the rising income share of the top percentile in the global distribution. We find that in 2005, individuals in the top 1% of the world had an annual average income of PPP$90,000. These individuals were on average 214 times richer than those in the poorest 21% of the world, who were living below PPP$1.25-a-day. Put differently, the richest 1% in the world, or 65 million people, had a total income equal to 10 times that of the poorest 21%, or 1.4 billion people.

Inequality in the global distribution can be decomposed into within-country and between-country components using decomposable inequality measures. We find that between-country inequality declined modestly during 1988–2005, whereas within-country inequality increased substantially. Between-country inequality is the larger component of global inequality, comprising 64–81% of overall inequality, depending on the inequality measure (MLD or Theil T) and year. Nevertheless, even if between-country inequality were eliminated, global inequality would remain at about the same level as in a high-inequality country such as China.

Although the extent to which global inequality has risen depends on the measure used, global poverty has declined substantially in recent decades. Given that inequality within countries has tended to rise, this decline has been driven by aggregate growth in low- and middle-income countries. It does not follow, however, that continued aggregate growth is the only way to continue to reduce global poverty. Redistribution of income within countries—by checking or reversing the rise in within-country inequality—could also make a significant contribution to the reduction of global poverty.

ACKNOWLEDGMENTS

We are very grateful to the editors, Tony Atkinson and François Bourguignon, for their detailed comments, and to Branko Milanovic for discussions regarding data. For helpful comments, we would also like to thank Angus Deaton and participants of the "Income Distribution Handbook Conference" at the Paris School of Economics in April 2013.

APPENDIX. ESTIMATES OF GLOBAL INEQUALITY BASED ON THE COMMON SAMPLE OVER TIME

The common sample over time for which we have both survey and PPP data in *all* five benchmark years comprises 67 countries. Table 11.A1 shows the total population of these 67 countries in each year and their share of the world population in that year.

Table 11.A1 Country and population coverage of common sample over time

Year	Number of countries	Population in billions (% of world population)
1988	67	4.16 (82)
1993	67	4.45 (80)
1998	67	4.76 (80)
2002	67	4.98 (80)
2005	67	5.13 (79)

Source: Authors' calculations.

Table 11.A2 Global inequality with and without top incomes, 1988–2005, for common sample over time

Year	Income share of top percentile (%)	Income share of top decile (%)	Gini	Between-country Gini (% of global Gini)	MLD	Between-country MLD (% of total)	Within-country MLD (% of total)	Theil T	Between-country Theil T (% of total)	Within-country Theil T (% of total)
With top incomes										
1988	17.7	58.8	0.730	0.653 (89)	1.159	0.913 (79)	0.246 (21)	1.130	0.796 (70)	0.334 (30)
1993	16.9	57.4	0.726	0.634 (87)	1.155	0.862 (75)	0.294 (25)	1.100	0.747 (68)	0.353 (32)
1998	18.9	59.1	0.722	0.632 (88)	1.093	0.796 (73)	0.297 (27)	1.140	0.754 (66)	0.386 (34)
2002	21.3	60.2	0.729	0.642 (88)	1.115	0.830 (74)	0.286 (26)	1.170	0.781 (67)	0.389 (33)
2005	21.5	58.3	0.721	0.624 (87)	1.141	0.801 (70)	0.340 (30)	1.153	0.722 (63)	0.431 (37)
Without top incomes										
1988	11.4	54.9	0.710	0.646 (91)	1.085	0.886 (82)	0.198 (18)	0.980	0.780 (80)	0.200 (20)
1993	11.6	54.7	0.706	0.630 (89)	1.083	0.843 (78)	0.239 (22)	0.962	0.738 (77)	0.223 (23)
1998	13.6	55.5	0.698	0.624 (89)	1.008	0.772 (77)	0.236 (23)	0.963	0.734 (76)	0.228 (24)
2002	13.3	56.0	0.705	0.633 (90)	1.029	0.798 (78)	0.231 (22)	0.991	0.760 (77)	0.231 (23)
2005	13.2	53.8	0.693	0.611 (88)	1.044	0.767 (73)	0.277 (27)	0.939	0.691 (74)	0.248 (26)

Source: Authors' calculations.

REFERENCES

Ackland, R., Dowrick, S., Freyens, B., 2004. Measuring Global Poverty: Why PPP Methods Matter. Paper presented at the 2004 Conference of the International Association for Research in Income and Wealth, Cork, Ireland, August. http://www.iariw.org/papers/2004/dowrick.pdf.

Almås, I., 2012. International income inequality: Measuring PPP bias by estimating Engel curves for food. Am. Econ. Rev 102 (1), 1093–1117.

Alvaredo, F., Londoño Vélez, J., 2013. High Incomes and Personal Taxation in a Developing Economy: Colombia 1993-2010, Commitment to Equity Working Paper 12.

Anand, S., 1983. Inequality and Poverty in Malaysia: Measurement and Decomposition. Oxford University Press, Oxford.

Anand, S., Kanbur, S.M.R., 1993. Inequality and development: a critique. J. Develop. Econ. 41 (1), 19–43.

Anand, S., Segal, P., 2008. What do we know about global income inequality? J. Econ. Lit. 46 (11), 57–94.

Anand, S., Sen, A., 2000. The income component of the human development index. J. Human Develop. 1 (1), 83–106.

Anand, S., Segal, P., Stiglitz, J.E., 2010. Introduction. In: Anand, S., Segal, P., Stiglitz, J.E. (Eds.), Debates on the Measurement of Global Poverty. Oxford University Press, Oxford.

Atkinson, A.B., 1970. On the measurement of inequality. J. Econ. Theory 2 (3), 244–263.

Atkinson, A.B., 2007. Measuring top incomes: methodological issues. In: Atkinson, A.B., Piketty, T. (Eds.), Top Incomes over the Twentieth Century: A Contrast between Continental European and English-Speaking Countries. Oxford University Press, Oxford.

Atkinson, A.B., Brandolini, A., 2001. Promise and pitfalls in the use of 'secondary' datasets: income inequality in OECD countries. J. Econ. Lit. 39 (3), 771–799.

Atkinson, A.B., Brandolini, A., 2010. On analyzing the world distribution of income. World Bank Econ. Rev. 24 (1), 1–37.

Atkinson, A.B., Piketty, T., Saez, E., 2011. Top incomes in the long run of history. J. Econ. Lit. 49 (1), 3–71.

Banerjee, A., Piketty, T., 2010. Top Indian incomes, 1922-2000. In: Atkinson, A.B., Piketty, T. (Eds.), Top Incomes: A Global Perspective. Oxford University Press, Oxford.

Beddoes, Z.M., 2012. For Richer, For Poorer. The Economist. 13 October 2012, http://www.economist.com/node/21564414.

Bernstein, A.R., 2011. Political cosmopolitanism. In: Chatterjee, D. (Ed.), Encyclopedia of Global Justice. Springer, London.

Berry, A., Bourguignon, F., Morrisson, C., 1983. The level of world inequality: how much can one say? Rev. Income Wealth 29 (3), 217–241.

Boltho, A., Toniolo, G., 1999. The assessment: the twentieth century—achievements, failures, lessons. Oxf. Rev. Econ. Pol. 15 (4), 1–17.

Bourguignon, F., 2011. A turning point in global inequality . . . and beyond. In: Krull, W. (Ed.), Research and Responsibility: Reflection on Our Common Future. CEP Europäische Verlaganstalt Leipzig Gmbh, Leipzig.

Bourguignon, F., Morrisson, C., 2002. Inequality among world citizens: 1820-1992. Am. Econ. Rev. 92 (4), 727–744.

Bourguignon, F., Levin, V., Rosenblatt, D., 2004. Declining international inequality and economic divergence: reviewing the evidence through different lenses. Écon. Intern. 100, 13–25.

Burkhauser, R.V., Feng, S., Jenkins, S.P., Larrimore, J., 2009. Recent Trends in Top Income Shares in the USA: Reconciling Estimates from March CPS and IRS Tax Return Data, NBER Working Paper 15320, Cambridge, MA.

Chen, S., Ravallion, M., 2001. How did the world's poorest fare in the 1990s? Rev. Income Wealth 47, 283–300.

Chen, S., Ravallion, M., 2008. The Developing World is Poorer than We Thought, but No Less Successful in the Fight against Poverty, World Bank Policy Research Working Paper 4703, Washington, DC.

Chen, S., Ravallion, M., 2010. China is poorer than we thought, but no less successful in the fight against poverty. In: Anand, S., Segal, P., Stiglitz, J.E. (Eds.), Debates on the Measurement of Global Poverty. Oxford University Press, Oxford, pp. 327–340.

Chen, S., Ravallion, M., 2012. An Update to the World Bank's Estimates of Consumption Poverty in the Developing World. Briefing note prepared by Shaohua Chen and Martin Ravallion, Development Research Group, World Bank, Washington, DC (03-01-12).
Chotikapanich, D., Valenzuela, R., Prasada Rao, D.S., 1997. Global and regional inequality in the distribution of income: estimation with limited and incomplete data. Emp. Econ. 22, 533–546.
Collier, P., 2007. The Bottom Billion: Why the Poorest Countries are Failing and What Can Be Done About It. Oxford University Press, Oxford.
Deaton, A., 2005. Measuring poverty in a growing world (or measuring growth in a poor world). Rev. Econ. Stat. 87 (1), 1–19.
Deaton, A., 2010. Price indexes, inequality, and the measurement of world poverty. Am. Econ. Rev. 100 (1), 5–34.
Deaton, A., Dupriez, O., 2011. Purchasing power parity exchange rates for the global poor. Am. Econ. J. Appl. Econ. 3 (2), 137–166.
Deaton, A., Heston, A., 2010. Understanding PPPs and PPP-based national accounts. Am. Econ. J. Macroecon. 2 (4), 1–35.
Deaton, A., Kozel, V. (Eds.), 2005. The Great Indian Poverty Debate. Macmillan, New Delhi.
Deininger, K., Squire, L., 1996. A new data set measuring income inequality. World Bank Econ. Rev. 10 (3), 565–591.
Dikhanov, Y., Ward, M., 2002. Evolution of the Global Distribution of Income, 1970–99. Paper prepared for the 53rd Session of the International Statistical Institute held in Seoul, Republic of Korea, August 22–29, 2001 and updated for the 5th Conference on Globalization, Growth and (In)Equality held in Warwick, England, March 15-17, 2002.
Dowrick, S., Akmal, M., 2005. Contradictory trends in global income inequality: a tale of two biases. Rev. Income Wealth 51 (2), 201–229.
Durlauf, S.N., Johnson, P.A., Temple, J.R.W., 2009. The econometrics of convergence. In: Mills, T.C., Patterson, K. (Eds.), Palgrave Handbook of Econometrics, Volume 2: Applied Econometrics. Palgrave Macmillan, New York, NY.
Firebaugh, G., 1999. Empirics of world income inequality. Am. J. of Sociol. 104 (6), 1597–1630.
Firebaugh, G., 2003. The New Geography of Global Income Inequality. Harvard University Press, Cambridge, MA.
Guio, A.-C., 2005. Income poverty and social exclusion in the EU25. In: Statistics in Focus, Population and Social Conditions. Eurostat, 13/2005.
Harttgen, K., Klasen, S., Vollmer, S., 2013. An African growth miracle? Or: what do asset indices tell us about trends in economic performance? Rev. Income Wealth 59 (Special Issue), S37–S61.
Jerven, M., 2013. Comparability of GDP estimates in sub-Saharan Africa: the effect of revisions in sources and methods since structural adjustment. Rev. Income Wealth 59 (Suppl. S1), S16–S36.
Kolm, S.-C., 1969. The optimal production of social justice. In: Margolis, J., Guitton, H. (Eds.), Public Economics. An Analysis of Public Production and Consumption and Their Relations to the Private Sectors. Macmillan, London.
Korzeniewicz, R., Moran, T., 1997. World-economic trends in the distribution of income, 1965–1992. Am. J. Sociol. 102 (4), 1000–1039.
Lagarde, C., 2013. A New Global Economy for a New Generation. Managing Director, International Monetary Fund, Speech as prepared for delivery, International Monetary Fund, Davos, Switzerland.
Lakner, C., Milanovic, B., 2013. Global Income Distribution: From the Fall of the Berlin Wall to the Great Recession. World Bank Policy Research Working Paper 6719, December, Washington, DC.
Melchior, A., Telle, K., Wiig, H., 2000. Globalisation and Inequality: World Income Distribution and Living Standards, 1960–1998. Royal Norwegian Ministry of Foreign Affairs, Studies on Foreign Policy Issues, Report 6B.
Milanovic, B., 1998. Income, Inequality and Poverty During the Transition from Planned to Market Economy. World Bank Regional and Sectoral Studies, Washington, DC.
Milanovic, B., 2002. True world income distribution, 1988 and 1993: first calculation based on household surveys alone. Econ. J. 112 (476), 51–92.

Milanovic, B., 2005. Worlds Apart: Measuring International and Global Inequality. Princeton University Press, Princeton.
Milanovic, B., 2012. Global inequality recalculated and updated: the effect of new PPP estimates on global inequality and 2005 estimates. J. Econ. Inequal. 10 (1), 1–18.
Nagel, T., 2005. The problem of global justice. Philos. Publ. Aff. 33, 113–147.
Neary, J.P., 2004. Rationalizing the Penn World Table: true multilateral indices for international comparisons of real income. Am. Econ. Rev. 94 (5), 1411–1428.
Pickett, K., Wilkinson, R., 2010. The Spirit Level: Why Greater Equality Makes Societies Stronger. Penguin, London.
Ravallion, M., 2004. Competing concepts of inequality in the globalization debate. In: Collins, S.M., Graham, C. (Eds.), Brookings Trade Forum 2004. Globalization, Poverty, and Inequality. Brookings Institution Press, Washington, DC.
Ravallion, M., 2009. Do Poorer Countries Have Less Capacity for Redistribution? World Bank Policy Research Working Paper 5046, Washington, DC.
Reddy, S., Pogge, T., 2010. How *not* to count the poor. In: Anand, S., Segal, P., Stiglitz, J.E. (Eds.), Debates on the Measurement of Global Poverty. Oxford University Press, Oxford.
Rodrik, D., 2013. The Past, Present, and Future of Economic Growth, Global Citizen Foundation. Working Paper 1, http://www.sss.ias.edu/files/pdfs/Rodrik/Research/GCF_Rodrik-working-paper-1_-6-24-13.pdf.
Sala-i-Martín, X., 2002. The Disturbing 'Rise' of Global Income Inequality, NBER Working Paper 8904, Cambridge, MA.
Sala-i-Martín, X., 2006. The world distribution of income: falling poverty and . . . convergence, period. Q. J. Econ. 121 (2), 351–397.
Schultz, T.P., 1998. Inequality in the distribution of personal income in the world: how it is changing and why. J. Popul. Econ. 11 (3), 307–344.
Seers, D., 1972. What are we trying to measure? J. Develop. Stud. 8 (3), 21–36.
Segal, P., 2008. Review of *The Bottom Billion*, by Paul Collier. Renewal 16 (3/4), 152–154.
Segal, P., 2011. Resource rents, redistribution, and halving global poverty: the resource dividend. World Dev. 39 (4), 475–489.
Sen, A., 1973. On Economic Inequality. Clarendon Press, Oxford.
Sen, A., 2000. Social justice and the distribution of income. In: Atkinson, A.B., Bourguignon, F. (Eds.), Handbook of Income Distribution, Volume 1. Elsevier, Amsterdam.
Sen, A., 2009. The Idea of Justice. Harvard University Press, Cambridge, MA.
Sumner, A., 2012. Where do the poor live? World Dev. 40 (5), 865–877.
Székely, M., Hilgert, M., 1999. What's Behind the Inequality We Measure: An Investigation Using Latin American Data, Inter-American Development Bank Working Paper #409, Washington, DC.
Tett, G., 2012. Income Inequality Comes Out of the Igloo at Davos. FT.com, 25 January 2012, http://blogs.ft.com/the-world/2012/01/income-inequality-comes-out-of-the-igloo-at-davos/.
World Bank, 1990. World Development Report 1990: Poverty. World Bank and Oxford University Press, Washington, DC.
World Bank, 2008. 2005 International Comparison Program: Tables of Final Results. World Bank, Washington, DC.
World Economic Forum, 2012. Global Risks 2012, seventh ed. World Economic Forum, Geneva.
Young, A., 2012. The African Growth Miracle, NBER Working Paper 18490, Cambridge, MA.

CHAPTER 12

Gender Inequality

Sophie Ponthieux*, Dominique Meurs†
*Institut national de la statistique et des études économiques, France
†University Paris Ouest Nanterre, EconomiX and Ined, France

Contents

Handbook of Income Distribution, Volume 2A
ISSN 1574-0056, http://dx.doi.org/10.1016/B978-0-444-59428-0.00013-8

Abstract

The chapter examines how the various dimensions of economic inequality between men and women are analyzed today. Beyond the gender wage gap—a central issue—and of course the still far from equal sharing of housework, the chapter also reviews research on gender inequality in access to self-employment, the gender gap in pensions, and the emerging topic of a gender gap in wealth, attempting to highlight the paths between the various facets of gender inequality. Throughout the review, much attention is paid to measurement issues, the scope of empirical evidence, and to limitations due either to the small number of large and comparable data sets or to conventional approaches that limit the possibilities to compare men and women.

Keywords

Gender, Inequality, Discrimination, Segregation, Household, Family, Labor force participation, Family-friendly policies, Income, Living standard, Wages, Self-employment, Pensions, Wealth, Time-use surveys, Unpaid work, Housework, Time allocation

JEL Classification Codes

B54, D13, D14, D19, D31, I32, J16, J21, J22, J24, J26, J31, J32, J38, J71, J78, H31, H55, Z13

12.1. INTRODUCTION

While it is undisputable that the economic status of men and women has greatly converged in the second half of the twentieth century, it is also well documented that it is still not equal: women receive, on average, lower employment incomes than men, they more often work part-time or not at all, and they carry out the lion's share of unpaid work in the home. But while there are so many clues suggesting that women's overall income is lower than that of men, there is no straightforward measure that would allow the size of such an overall gender gap in income or in economic well-being to be assessed. The main limitation is that, in most income data sets, only some components of income, essentially related to work, are received and available at an individual level; the others are either received by households or measured as if received at the household level. The rationale for this household-level conceptualization is that multiperson households are assumed to benefit equally from shared resources within the household. This strong assumption about the distribution of income is made in the absence of reliable indications of the actual extent of intrahousehold sharing, so indicators of gender inequality in income or living standards are limited, if not biased. Then, as long as it is not possible to measure individual income—that is, being able to distribute the household income between the household members—gender inequalities are best assessed by various outcomes such as wages, pensions, or time spent in unpaid work than on the basis of a synthetic indicator assuming that all individuals who live together necessarily achieve an equal level of economic well-being.

The household is not only a measurement issue: many men and women live together in couple households, and many decisions are made in the context of the household. Then, what happens within the household, especially regarding the allocation of time to paid and unpaid work, is central to the understanding of the gender gap in economic outcomes, particularly gender differentials in earnings, pensions, and wealth accumulation. So it is not surprising that the issues of childcare and housework are repeatedly put forward in the analysis of various facets of gender inequality. Labor market institutions, firms' management, public policies, and social norms, however, also play a major role in shaping male and female behavior, constraints, and opportunities. Conversely, labor market status, earnings, and career perspectives influence the allocation of time and relative bargaining power within the household.

In other words, studying gender inequality entails considering the interaction of several influences relating to the family as well as the public sphere, as tentatively depicted in Figure 12.1. It also brings together research questions otherwise mostly addressed in separate strands of literature.

The aim of this chapter is to provide an up-to-date review of what is known and what we do not know, at least not well, about gender economic inequalities. The review is centered on "Western"/industrialized countries. This is not because other countries would be less interesting or less prone to gender inequality—it may even take more radical forms—but because it would lead to include other (too) vast issues, especially those of economic development and economic transition.

The chapter is organized as follows: Section 12.2 provides a brief overview of the current standard approach to income and living standards in statistical information, in

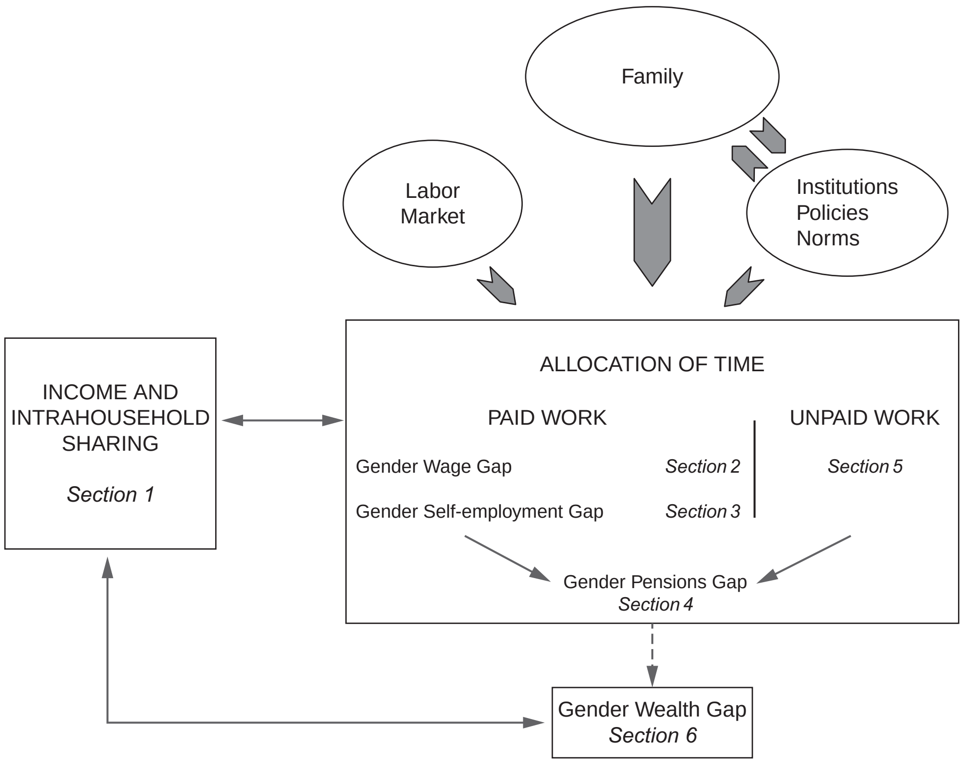


Figure 12.1 Main dimensions of gender economic inequality.

which the basic unit is the household, and its consequences for the measurement and analysis of income inequality between men and women. Sections 12.3–12.5 present an overview of the gender earnings gap. Section 12.3 is devoted to the central issue of the gender gap in wages, which has been and remains abundantly researched. Section 12.4 turns to self-employment, focusing on the question of the underrepresentation of women in this status. The gender gap in pensions, a result of past career and family trajectories, is examined in Section 12.5. Section 12.6 addresses the questions of "time and money" in the household, focusing on the unbalanced distribution of paid and unpaid work between men and women. Section 12.7 reviews the relatively small but growing literature on gender inequality in wealth.

12.2. INDIVIDUAL AND/OR HOUSEHOLD INCOME AND LIVING STANDARDS: FROM MEASUREMENT ISSUES TO CONCEPTUAL ISSUES AND BACK TO MEASUREMENT ISSUES

How to assess the extent of gender inequality in economic well-being? In rich countries, the most usual indicator of economic well-being is the living standard, a notion based on income. But in income statistics, many income components are available only for households,[1] not individuals, and living standards—that is, the household income adjusted for the household size and composition—are measured assuming that all incomes are shared within households and are assumed to be equal for all members of a given household. In other words, no inequality can be found between men and women living together in a couple-household, a very frequent situation. Assessing a gender gap in economic resources or well-being on such bases is hardly possible: If incomes are not pooled and fully shared, using household-level information may result in seriously biased estimates of income and living standard inequality. But how seriously? The problem is that not much is known about the actual distribution of income within households. This first section, then, is not so much about gender inequality in income as about the limitations to its measurement, the methodology and conceptualization behind measurement, and the implications for the analysis of gender inequality.

That income statistics do not systematically provide individual-level income components is in accordance with the current international standard, following the guidelines

[1] A household is defined as "either (a) a person living alone [. . .] or (b) a group of two or more persons who combine to occupy the whole or part of a housing unit and to provide themselves with food and possibly other essentials for living. The group may be composed of related persons only or of unrelated persons or of a combination of both. The group may also pool their income" (Canberra Group, 2011, p. 25). Note that there is not necessarily a strict correspondence between the dwelling, the household, the family, and the fiscal unit. There are also many debates on the actual extent of a "household," especially in relation to the increasing share of reconstituted or multiresidence families (e.g., children of divorced or separated parents), further complicating household income statistics.

defined by the Canberra Group[2] (2001, 2011). Acknowledging that economic well-being is more an individual than a collective notion, these guidelines nevertheless refer to the household as the best statistical unit for the production of income statistics, for pragmatic reasons: "The starting unit is the individual, but as individuals typically share income with the other persons with whom they live, most surveys collect information on the income streams of all members of a larger statistical unit, most commonly the household. [. . .] A full appraisal of income sharing within a household would require collecting data on the income transfers made within the household which would obviously be very difficult to implement. For these reasons, the choice of the household as the basic data collection unit for collecting income data remains the best compromise" (Canberra Group, 2011, pp. 24–25).

In support of pragmatism, one must recognize that it is really very difficult to "distribute" all the income components between individuals. First, some incomes are difficult to attribute precisely to one or another household member—for instance, family benefits or the income from assets owned jointly by spouses. Second, the household members may actually share all or part of their incomes and benefit together from shared assets such as housing; measuring an individual's income would require knowing the amount of income he/she receives from (or transfers toward) another individual within the household. In addition, living with others results in economies of scale (from sharing a dwelling, equipment, etc.); although this is not, strictly speaking, "income," it must also be taken into account in the comparison of living standards between individuals. So, except in the case of one-person households, measuring income or living standards at the individual level requires either restricting the notion of income to components that can be precisely attributed to one or another household member or having some knowledge of the extent of income sharing and the distribution of income within households. The same goes for wealth. But knowledge is replaced by two crucial, and distinct, assumptions: that of income pooling and that of equal living standards in the household—as if equality was an automatic consequence of income pooling, while it is an independent assumption. Lack of knowledge replaced by assumptions is the main

[2] The Canberra Group is one of the United Nations' groups of experts on statistical methodology, known as "city groups" (http://unstats.un.org/UNSD/methods/citygroup/index.htm). The Canberra Group was established at the initiative of the Australian statistical agency in 1996, bringing together representatives of about 30 national statistical agencies and international organizations such as the Organisation for Economic Co-operation and Development, the International Labour Organization, and the World Bank. Its purpose was to define a harmonized framework for income statistics, both consistent with the concepts of national accounts and allowing pertinent cross-country comparisons at a micro level. It resulted in a first edition of guidelines (Canberra Group, 2001), which have become an international reference (see also Smeeding and Weinberg, 2001). The concepts and definitions remain broadly the same in the 2011 update. "Income" is broadly understood as the amount a person/household can spend or save over a given period of time without being worse off at the end of this period than at the beginning.

obstacle to the comparison of income and living standards between individuals in general and between men and women in particular.

This section starts with a brief presentation of the components of household income, focusing on the articulation between categories of incomes and income units and of the standard approach to living standards. Then it turns to the conceptual background, presenting the unitary model of the household and its limits and an overview of other, non-unitary models and of the socioeconomic approach to the financial organization of the household from the perspective of their implications for operational statistical practice. The consequences of lacking knowledge and information at the individual level, especially the consequences for the analysis of gender inequality, are discussed, focusing on the assessment of poverty and the implications of a "unitary" perspective in policies against poverty.

12.2.1 Measuring Income: Components and Units

Household disposable income—that is, the amount a household can spend or save without drawing on its assets—is measured as the sum of all the incomes received in the household over a given period of time net of the sum of social contributions and direct taxes paid by the household and transfers to other households. Three main types of income can be received in a household: income from work/employment, income from property, and transfers, some received from other households (e.g., alimonies, child support) and most received from the state.[3] In this set of components, the primary unit of income is more or less obvious: income from work and employment, including wages, salaries, profits or losses from self-employment, some state transfers related to work (unemployment, sickness and maternity benefits, pensions), some other social benefits such as incapacity benefits or scholarships, and some transfers with other households, such as alimonies, are clearly received by individuals. As for deductions, the social contributions attached to earnings or pensions are "individual," too. But some other components can be individual or "collective": for instance, who receives the income from property depends on who the owner is. Attributing this income to one or another member of the household then requires detailed information on ownership, that is, on who holds the assets or, in the case of joint ownership, the shares held by the different owners. Yet

[3] The Canberra Group also mentions the value of goods and services produced by the household for its own consumption as well as social transfers in kind. These elements, while not monetary incomes, increase the household's actual resources. However, measuring the goods and services produced for the household's own consumption or distributing the value of state transfers in kind between households raises serious difficulties of estimation (especially a very high sensitivity to estimation assumptions), and the current definition of household disposable income does not yet include them (even though some data sets provide estimates, for example, imputed rent). This chapter will not go further into this aspect of the measurement of income, which involves long-running debates on the concept of income, recently renewed with the "Stiglitz-Sen-Fitoussi" report (Stiglitz et al., 2009).

wealth is perhaps even more conceptualized at household level than income; we return to this issue later in the chapter (Section 12.7). Taxation can be joint or separate. Some other components are more "collective," for instance, family or housing benefits. Of course, none of these doubts arise in the case of one-person households, for which the household and the individual income are the same.

Identifying the unit who receives/pays is one thing, but what counts in practice is the level at which the information is available in statistical data sets; by definition, strictly household-level data sets do not provide any information at the individual level. Individual datasets most often provide information limited to earnings (sometimes pensions and social benefits); and data sets providing both individual-level and household-level information do not always detail all the components collected at an individual level.[4] Table 12.1 displays a tentative classification of the components of household income by primary income unit and the level at which they are most often available in large-scale datasets (excluding occasional surveys). It shows clearly that, on the basis of the information available in data sets, employment-related income is the only notion of income that can be easily implemented at the individual level. In consequence, the statistical information on income currently available in most large-scale data sets does not allow one to compute a disposable income at the individual level if the individual does not live alone.

12.2.2 From the Household's Disposable Income to the Standard of Living at the Individual Level: The Statistical Approach

The statistical notion of living standard (or "equivalent income" or "income per consumption unit") is intended to make comparable the economic well-being of households of different sizes and composition. Its quality as a proxy for economic well-being is much debated (see also Decancq et al., 2014, Chapter 2, in this volume), but for the time being, it remains the most often used and the basis for the measurement of poverty thresholds (or, in the United States, poverty lines).

The standard of living is a construction based on the disposable income (or consumption) of a one-person household taken as a measure of his or her economic well-being. When the household counts more than one person, it is measured so as to take into account the fact that the needs of two (three, etc.) individuals living together are less than twice (three times, etc.) those of one person living alone because of the economies of

[4] For instance, European Union Statistics on Income and Living Conditions (EU-SILC), the European statistical source on income and living conditions, which is the basis for the computation of a system of indicators, provides information at both levels but (at the time this chapter was written) does not give the details of certain individual benefits (maternity benefits or parental leave allowance, for instance, are not detailed at the individual level and are aggregated with family benefits at the household level), nor does it provide the detailed amount of taxes when the household is not a married couple or when taxation is separate. In addition, it aggregates taxes and social contributions at the household level.

Table 12.1 Income components, income units, and availability in statistical data sets

Category	Income component	Income unit	Availability in data sets
Primary distribution	*1. Income from employment*		
Market income	a. Employee income (wages, salaries, bonuses, severance/termination pay)	Individual	Individual level Household level
	b. Income from self-employment (profit/loss from unincorporated enterprise)	Individual[a]	Individual level Household level
	2. Property income		
	c. Income from financial assets (interest, dividends)	Owner (individual or household)	Household level
	d. Income from nonfinancial assets (rent)		
	e. Royalties		
Redistribution	*3. Current transfers received*		
State transfers	f. Social security pensions/schemes	Individual	Individual level
	g. Other insurance benefits (unemployment, sickness, incapacity, etc.)	Individual	Individual level Household level
	h. Family, housing benefits	Household	Household level
	i. Other social benefits (universal or means-tested)	Individual or household	Household level
	j. Current transfers from other households (alimony, child or parental support)	Individual or household	Household level
	4. Current transfers paid		
Taxes and contributions	k. Social insurance contributions	Individual	Individual level Household level
	l. Direct taxes	Fiscal unit (individual or household)	Household level
	m. Current interhousehold transfers paid	Individual or household	Household level
	Disposable income = 1 + 2 + 3 − 4	Household	

[a]The "individual" nature of incomes from self-employment is questionable in the case of a family-owned business. There are, in general, well-known difficulties in measuring the income from self-employment (see Canberra Group, 2011, pp. 34–35). One particularity is that self-employment income includes income from both employment and capital; the income it yields is generally measured as the amount of profit over an accounting period (or losses; it can then be negative). Another particularity is that self-employment includes a specific category of employment: "family workers," who participate in the business of a family member but are not paid.

scale that result from sharing a dwelling and durable and consumption goods and the benefits of the household's production. The additional income needed to keep the household at the same level of economic well-being when additional members are included in the household is difficult to measure because individuals' consumption within households is not observed. It is most often estimated on the basis of the observed expenditure of households of different sizes and demographic composition. These estimations result in "equivalence scales," which give a weight, assumed to reflect the additional income needed relative to a one-person household, to each additional household member. Whatever the equivalence scale actually implemented, the weight of any additional member is less than 1 (because, as mentioned above, adding a second person to a one-person household does not double the needs of the household). The dominant statistical approach to the standard of living (or equivalent income) currently uses the so-called "OECD-modified" equivalence scale,[5] which gives a weight of 0.5 to an additional adult in the household and a weight of 0.3 to an additional child (a child being an individual younger than 14 years old).

While in this section we are more interested in the assumptions behind the measurement of the household's standard of living and its meaning at the individual level, it is worth briefly illustrating the difference between the household disposable income (INC), a "per head" approach (INC_{PH}), and the standard of living, or "equivalent" disposable income (INC_{EQ}). There is, of course, no difference for a one-person household: in this case $INC_{PH} = INC_{EQ} = INC/1$. If the household is composed of two adults, then $INC_{PH} = INC/2$ and $INC_{EQ} = INC/(1+0.5)$; if it is a couple with one child, $INC_{PH} = INC/3$ and $INC_{EQ} = INC/(1+0.5+0.3)$; INC_{EQ} is always greater than INC_{PH}; the difference accounts for economies of scale. Leaving aside a possible debate on the weightings, it is reasonable (and widely accepted) that INC_{EQ} is a better basis than INC or INC_{PH} for comparing the level of economic well-being between these three households. But current statistical practice goes a step further, since each individual in a given household is considered to achieve the level of economic well-being he or she would achieve if living alone with an income equal to the household's equivalent income; in other words, all the household's members have the same living standard—that of the household. As pointed out by Woolley and Marshall (1994): "The standard approach solves the problem of measuring resource distribution within households by ignoring it" (p. 429). In turn, this practice raises many questions about the actual meaning

[5] The "OECD-modified scale" has gradually replaced the "OECD-scale" or "Oxford scale" (which gave a weight of 0.7 to an additional adult and 0.5 to an additional child). Neither is unanimously validated, and there are various other possible weightings. The many debates on the estimation of equivalence scales will not be recounted here (see, e.g., Lechene, 1993; Van Praag and Warnaar, 1997; a survey in Lewbel and Pendakur, 2008; a discussion of statistical practice in OECD, 2013). Note that American statistics do not explicitly refer to an equivalence scale, but the official poverty lines (measured for households of different sizes/composition and in different locations) are implicit equivalence scales.

of indicators *at the individual level* (including indicators of poverty) based on the household's equivalent income: if one of the household members holds back some or all of his or her income from the common pool, or if the pooled income is not equally distributed among the household members (or if the equivalence scale does not allow for economies of scale to be "distributed" at individual level; see also Browning et al., 2006a), the approach is much less relevant. This highlights how essential the assumptions of income pooling *and* equal sharing within the household are both necessary conditions for using equivalence scales (Lise and Seitz, 2011) and deriving individual-level indicators from variables measured at the household level. The standard approach also results in a measure of individual well-being, which, by construction, ignores the possibility of inequality within the household and, by construction, makes intrahousehold inequality virtually impossible to assess.

12.2.3 Behind the Statistical Approach: The Household "As If" an Individual

Curiously enough, the statistical practice of attributing the same living standard to all the individuals living in a household is rooted in a conceptualization of the household "as if" it was a single individual. The household "as an individual," first developed in the framework of consumer theory, is the household of the unitary model. This model has many implications for the analysis of gender inequality (Section 12.4); the focus here is on its assumptions and limitations in relation to the practice of income measurement.

12.2.3.1 The Unitary Model of Household Behavior

The unitary model consists in a transposition of individual (consumer) behavior to the household level: according to its preferences, the household maximizes a single utility function under a single budget constraint. For a household to function "as if" an individual, two main assumptions are needed.

First, individual preferences must converge, one way or another, so that the household can be considered a single decision unit. Considering that the actual agent of the consumer theory is a household (a family), Samuelson (1956) proposed that the family acts as if it were maximizing a joint consensual welfare function, a solution that does not explain the process resulting in a consensus. Another solution to the shift from several individual preferences to a single utility function was provided by Becker (1974a, 1991), proposing that the household is ruled by a "head"—an altruistic or "benevolent" member of the household—who cares about the welfare of the other family members. This would be consistent with the view of marriage based on mutual interest (the gains it provides compared to remaining unmarried) plus love and caring—the addition of optimal associations resulting from the marriage market even making the possibility of shared preferences plausible (Becker, 1973, 1974b). In this approach, the head of household transfers resources to the other household members; the household then acts as a single unit, since the only way the household members can increase their utility is to act in a way that maximizes the utility of the head of household, who in turn will increase his transfers.[6] However, this approach has been shown to hold only under very restrictive conditions (Bergstrom, 1989); a further assumption, which is actually crucial, is also necessary: the altruistic head of household has to be able to control the distribution of resources, meaning either he is richer than the other household members (Ben-Porath, 1982) or he has more power by other means (Folbre, 1986; Pollak, 1985).

Second, the incomes of the household members have to be fully pooled—a necessary condition for there to be only one budget constraint. Income pooling means that the way the income is used (by the household) depends only on the level of the whole/pooled income (and the household's preferences) and not on the origin of the income.

In the perspective of income measurement, the unitary model then tells us that there is no need to complicate income statistics with information collected at the individual level because it is only the total income of the household, not whose income it is, that counts in the end. However, the head of household does not necessarily treat all the household members equally; then the statistical practice of imputing the same living standard to all the household members may carry the approach a step further.

12.2.3.2 Methodological and Empirical Issues

The unitary model has been challenged on both theoretical and empirical grounds since the 1980s. At the theoretical level, the approach contradicts the methodological principle of individualism at the basis of neoclassical theory (Apps and Rees, 1988; Chiappori, 1992). According to this principle, there is no such thing as "group behavior," only outcomes from individuals' decisions. Folbre (1986) also underlines the paradox of "individuals who are entirely selfish in the market (. . .) [and] entirely 'selfless' within the family" (p. 247). The unitary model, along with Becker's theories on marriage and specialization within the family, also have been much discussed by feminist economists in a general examination of its implications for the analysis of gender (in)equality.[7]

Empirical results do not provide much support. Basically, a test of the unitary model consists of verifying that a change in the distribution of nonlabor income within the household does not modify the structure of the household's consumption or labor supply behavior (since both are expected to result only from the household's preferences and the budget constraint, i.e., the pooled income and prices); in short, how the (pooled) income is used should be independent from whose income it is. On the contrary, it has been

[6] Note that the "head" is not explicitly a "he": Becker's formulation here is "gender neutral." However, it is most likely "he" in the Beckerian approach to specialization within the family (see Section 12.6.3), as well as empirically, given the likelihood of women having, on average, lower independent incomes than their male partners.

[7] Nelson (1994), Bergmann (1995), Katz (1997), and Woolley (1996). See also an in-depth review of Becker's main assumptions founding his approach to the economics of the family by Pollak (2003).

shown that changes in the personal income level of the partners influence the household's allocations (Bourguignon et al., 1993; Browning et al., 1994; Phipps and Burton, 1998; Thomas, 1990). Empirical work on developing countries also casts doubt on the validity of the unitary approach (e.g., Kanbur and Haddad, 1994) to the point that, in the mid-1990s, several economists thought that the time had come "to shift the burden of proof" onto the unitary model (Alderman et al., 1995). The early results were supported by results obtained from the analysis of the impact of exogenous changes in the household members' relative incomes, which makes it possible to avoid possible endogeneity biases (for instance, a change in relative incomes resulting from a change in the labor supply). Here, an emblematic study is that of Lundberg et al. (1997); they show that a reform in the payment of family benefits in the United Kingdom—initially paid to fathers, then after the reform to mothers—resulted in an increase in household expenditure on children's clothes, a result not consistent with the unitary model.[8] Such conditions of "natural experiments" (more precisely, changes that exogenously redistribute the incomes within households) are difficult to find, and comparable studies do not exist for rich countries.[9] However, considering the results of empirical studies over a period running from 1994 to 2008 in different contexts and for outcomes as different as, for example, labor supply, children's health, savings, demand for clothes, and demand for alcohol and tobacco, Browning et al. (2011) concluded that: "the evidence seems overwhelming: a principal implication of the unitary model is rejected on a wide set of data sets for a wide range of outcomes" (p. 225).

12.2.4 Other Representations of the Household

Several other nonunitary models of household economic behavior have developed since the early 1980s. Parallel with these developments, a strand of research in sociology and economic psychology examines the financial organization of the household. We now look at these alternative economic models[10] and socioeconomic approaches and the alternatives for statistical practice they offer.

[8] Ward-Batts (2008) obtains the same result using a different type of data.

[9] Duflo (2003), who used a reform extending the benefit of a social pension program to a formerly excluded black population in South Africa, or Attanasio and Lechene (2002), who used the Mexican welfare program Progresa, in which transfers go exclusively to women, obtain results rejecting the income pooling assumption, and both suggest a shift of power in favor of women. Comparing Progresa and Procampo, another welfare scheme targeting men (farmers), Davis et al. (2002) found that the former results in better outcomes than the latter in terms of schooling expenditures and health, but the outcome does not differ in terms of households' total expenditure and food security (see also Ruiz-Arranz et al., 2006). Analyzing the impact of a law extending alimony rights to cohabitants in Brazil, Rangel (2006) also founds a shift of power toward women in decision making on time allocation and a redistribution of resources toward schooling expenditure on first-born girls.

[10] See Browning et al. (2011) for an in-depth review of models of household behavior and recent developments, and see Donni and Chiappori (2011) for a survey.

12.2.4.1 Nonunitary Models of the Household

Contrary to the unitary model, nonunitary models explicitly consider the individuals within the household in accordance with the methodological principle of individualism, and they take into account a notion of relative power in decision making, allowing for a gendered approach to intrahousehold organization. Nonunitary models generally consider two decision makers (spouses), each with his or her own utility function,[11] allowing for externalities—that is, including the partner's preferences in their utility functions. The partners' relative power in the decision-making process is taken into account either as the utility each partner would obtain outside the marriage or in a sharing rule. Two categories of goods are considered: private goods, corresponding to goods that can be consumed only by one partner (and will be included only in one partner's utility function), and public goods that can be consumed at the same time by both partners. Children do not directly influence household decisions, but their well-being is taken into account as a public good in the parents' preferences. Incomes are not necessarily assumed to be pooled (but they can be), and the origin of income contributes to the partners' relative power, as well as various other distribution factors, that is, factors that may influence the decision process but do not change the budget constraint: typically the partners' nonlabor incomes, education levels, health status, wealth, or social capital. The distribution factors can also be external to the partners' characteristics: the state of the labor market or the marriage market may shift the power balance through the opportunities they offer outside the partnership. Folbre (1997) also refers to "gender-specific environmental parameters" such as women's rights in general, marriage and divorce laws, and property rights. Agarwal (1997) extends the understanding to the influence that this environment, including social norms, can have on the actual scope for bargaining within the household.

Beyond these broad characteristics, nonunitary models differ in the assumptions made about whether decisions result in Pareto-efficient outcomes—that is, when improving one partner's well-being is not possible without reducing the other's—and in the way they represent the process of decision. There are three main categories of nonunitary models: cooperative bargaining models (Manser and Brown, 1980; McElroy and Horney, 1981), noncooperative bargaining models (Ulph, 1988; Woolley, 1988), and collective models (Bourguignon and Chiappori, 1992; Browning et al., 1994; Chiappori, 1988). Drawing on game theory, bargaining models have in common the assumption of a certain process of decision making but not necessarily that the outcome is Pareto-efficient; on the contrary, collective models do not specify a particular decision-making process but assume Pareto-efficiency, justified by the stability of the relationship (see more detailed presentations in Chiappori and Meghir, 2014, Chapter 16; Browning

[11] Considering two decision makers with their own preferences avoids the methodological problem of the unitary model; however, some level of "collective" rationality needs be introduced for the outcome to be Pareto-efficient (see Bourguignon, 1984).

et al., 2006b; Donni and Chiappori, 2011). However, Pareto-efficiency becomes difficult to assume when the household's environment is not stable, for example, when decisions change the partners' relative bargaining power or if what seem to be optimal in a given context can lead to nonoptimal outcomes when the context is changed. The exploration of these dynamic aspects is one of the directions taken by most recent research (e.g., Basu, 2006; see also Chiappori and Meghir, 2014, Chapter 16).

Bargaining models have introduced the notion of "threat point"; the basic idea is that neither partner would agree on a decision resulting in a lower level of individual utility than that he or she could achieve if the partners did not come to an agreement. The individual level of utility (the utility at the threat point) defines a partner's bargaining power. The threat point was formulated first as a divorce threat—a rather radical argument of negotiation, as underlined by Bergstrom (1996). Another formulation is that of the model of "separate spheres" (Lundberg and Pollak, 1993); here, the threat is not defined in terms of union dissolution but as minimal cooperation, with each partner being responsible for a given sphere (defined by social gender norms) of common consumption.[12,13] The interest of the bargaining approach is to provide a framework for analyzing unequal power relations within the household and how they can be influenced by environmental factors. Folbre (1986) emphasizes the importance of this difference from the unitary model: "It is the juxtaposition of women's lack of economic power with the unequal allocation of household resources that lends the bargaining power approach much of its persuasive appeal" (p. 251). However, these models present serious difficulties of estimation, especially the need to identify a threat point, as described by Himmelweit et al. (2013).

Because they do not refer to a specific decision process, collective models do not face this particular problem. They can be seen as a general class of cooperative models and are nowadays the dominant approach to household decision making. In these models, the partners maximize a household utility function, which can be seen as a weighted average of each partner's utility function; the weighting corresponds to a sharing rule, which depends on distributional factors and reflects the partners' relative power (instead of "points" where marriage or cooperation breaks down). Empirical results regularly show that the sharing rule is influenced, as expected, by various characteristics of the household members and contextual parameters (Browning et al., 2011). Empirical applications, however, often require more assumptions because neither the outcome nor the sharing rule are observable and statistical information is most often available only at household

[12] The difference from the external threat, where the bargaining power depends on the utility the partners would get outside the partnership, is that in Lunderg and Pollak's (1993) setting it depends on the resources the partners bring to the marriage. These two views of the threat point may lead to completely opposite policy implications.

[13] Other models that are not detailed here introduce the possibility of transfers between partners (Carter and Katz, 1993; Chen and Woolley, 2001) or specify several threat points (Bergstrom, 1996).

level. In addition, while variations in the sharing rule can be estimated, it is much more complicated to obtain estimations of the sharing rule itself[14] (Himmelweit et al., 2013).

Nonunitary models provide a conceptual alternative to the unitary approach and have allowed many improvements in the analysis of the household members' behavior within the household, but, in the perspective of statistical practice, they do not provide the operational alternatives that would allow going beyond the convenient assumptions of the unitary ground, in part because their estimation requires additional assumptions that reduce their scope and in part because of the lack of large-scale individual data on income.

12.2.4.2 Intrahousehold Finances: A Socioeconomic Perspective on Income in the Household

In sociology and economic psychology, the household (actually the couple-household, as in most economic models) is analyzed as a place of bargaining between partners of unequal power and may have conflicting interests. The household's domestic organization is explained within the main framework of the theory of resource exchange, which predicts that the partner with the highest resources—income, education, status—will have more power within the household (Blood and Wolfe, 1960; Sabatelli and Shehan, 1993). The notion of resources may include opportunities outside the relationship and the costs of an exit solution—an idea interestingly close to that of economic nonunitary models.

The issue of intrahousehold distribution of income is addressed in terms of financial organization, characterized by the management of and control over money by the partners. Another notion, not present in economics, is that of different categories of money—or "monies," according to the term introduced by Zelizer (1989, 1994)—depending on whose money it is and linking it to specific uses, some of which are shaped by gender norms: as in the case of labor division, money management reveals gendered relations of power (Tichenor, 1999). "Money" is not "income," but the questions addressed about money in the household are strikingly similar to those addressed in economics about the distribution of income within the household: Does the origin of money matter? Is it possible to lack money in a rich household? Who has power over what money? Who is responsible for given expenditures? Who makes decisions about what?

A very influential classification of money management systems, intended to reflect gradations of control over money, was proposed by Pahl (1983, 1989). She defined four main systems.[15] (1) The "whole wage" system: one partner hands over all his or her wage (minus an amount of "pocket money") to the other, who is in charge of managing the household's finance; this other partner may or may not have a personal income. (2) The "housekeeping allowance" system: here, one partner hands over an amount to be used for

[14] Examples of such estimations can be found in Couprie (2007) or Kalugina et al. (2009), who use methodologies that are clearly not likely to be implemented on a systematic basis.

[15] Further refinements distinguish between men's and women's control over resources (Pahl, 1995).

housekeeping spending and manages the rest him/herself. In these first two systems, only one partner has control over the household money. (3) The "pooling" or shared system: both partners manage and use the money as they need either for common or personal expenditure. (4) The "independent management" system: each partner keeps separate control over their income; unlike the three other systems, there is no notion of household money and neither partner has access to the other's money. This classification has been central in many dedicated studies of quantitative sociology and socioeconomics over the 1990s (e.g., Burgoyne, 1990; Burgoyne and Morison, 1997; Treas, 1993; Vogler, 1998) and remains the reference in recent empirical work (see Bennett, 2013).

One of the interests of Pahl's classification is that it provides usable survey questions[16]: typically, the respondents are asked whether they pool all, some, or none of the income, and sometimes the share of their income they pay into a "common pot" or the share they spend on their own private consumption.[17] Empirical work generally finds that the pooled system is more often observed in the case of married couples (Heimdal and Houseknecht, 2003; Ludwig-Mayerhofer et al., 2006; Lyngstad et al., 2011), couples who have children (Hamplova and Le Bourdais, 2009; Kenney, 2006; Laporte and Schellenberg, 2011; Vogler et al., 2008), in one-earner families or when the income imbalance is significant (Elizabeth, 2001; Halleröd, 2005), and when money is tight. It is less frequent when the partners have higher education levels (Laporte and Schellenberg, 2011), in the case of past partnerships, or when there are financial ties with other households (Burgoyne and Morison, 1997; Heimdal and Houseknecht, 2003; Treas, 1993). In summary, bearing in mind that different methodological options[18] and data limitations sometimes make the results difficult to compare, duration, the presence of public goods (especially children), a traditional division of labor, and the need to monitor low resources seem to have a positive influence on the probability of pooling

[16] Pahl's classification was based on a very small number (from a statistical point of view) of in-depth interviews—about 120 in all.

[17] Examples of such questionnaires can be found in the EU-SILC 2010 module on "Intra-household allocation of resources" (Eurostat, 2010), in the Statistics Canada General Social Survey 2007 (see Laporte and Schellenberg, 2011), or the International Social Survey Program (waves 1994 and 2002). An earlier example can be found in work by Woolley and Marshall (1994), who use Pahl's categories to compare various measures of intrahousehold inequality. In other formulations, the respondents are asked whether they see their earnings as their own or as the family's money and whether they feel free to draw on the common pool without asking permission. Questions on the respondents' opinion about or satisfaction with the household income also have been developed; this type of information was used in the estimation of the sharing rule by Kalugina et al. (2009) and Bonke and Browning (2009). Some studies use the existence of a joint bank account rather than questions on the financial regime (Woolley, 2003), but the two approaches are not comparable, as shown by Burgoyne et al. (2007).

[18] Options include whether to aggregate Pahl's categories, especially those of partial pooling and separate money. Aggregation is often intended to contrast collective systems—allocative and pooled systems—with all the other systems that are categorized as individualized (Vogler et al., 2008).

incomes. Beyond this general pattern, comparative studies show significant cross-country variation that might relate to different institutional and cultural contexts (Hamplova and Le Bourdais, 2009; Heikel et al., 2010; Heimdal and Houseknecht, 2003; Yodanis and Lauer, 2007). As for the overall incidence of the pooled system, the main result is that significant proportions of couples declare that they keep at least some of their money separate,[19] another dent in the unitary model insofar as the concept is the same as that of income pooling in the economic literature.

From the perspective of statistics appropriate for the study of income inequality between individuals, implementing "pooling" questions on a systematic basis has some appeal. It would be an easy way to determine whether incomes are fully pooled. If so, this would justify a unitary approach; if not, some additional questions on the share of different incomes held back from the "common pot" by the household members would make it possible to estimate the amount of income actually available to, or controlled by, each member. However, several questions can be raised. One is what is actually captured by the answers to pooling questions: a perception of more or less collective money, a practice of management, or actual control over one's own or the household's money? It is also difficult to deduce whether there is equality within the household from the pooling regime. Transfers between partners can take place without income pooling, and hence no pooling does not mean no sharing; conversely, income pooling does not necessarily entail equal or fair sharing. Another issue is how the respondents understand the income/money questions: contrary to in-depth interviews, where the researcher can interact with the respondent to ensure a common understanding, survey questions may be understood differently from their intended meaning. Finally, there is the question of the identity of the respondent: is it a unique respondent, representative of the household, or both partners? Questions asked of only one partner may provide reductive information, but asking both partners may lead to dealing with conflicting answers.[20,21] Beyond "technical" questions, the pertinence of the information obtained from survey questions on the pooling regime is questioned in qualitative research. For instance, exploring the case of partial pooling and separate finances, Ashby and Burgoyne (2008) suggest that they may conceal more complex practices relating to perceived ownership and commitment in the

[19] A presentation of the main results of the EU-SILC 2010 module on "Intra-household allocation of resources" shows that, among the 27 EU member states, about 70% of the households (with at least two adults)—ranging from as low as 50% in Austria to about 85% in Malta—reported that all the incomes are treated as common (Eurostat, 2012).

[20] In an analysis of individual or household responses to questions on perceived economic well-being, Plug and Van Praag (1998) found that, when asked about the minimum income needed to maintain a family like their own, there was no difference between the partners' responses in one-earner and two-earner couples, as long as the earnings differential was not substantial.

[21] The analysis of the Eurostat module on intrahousehold sharing of resources (EU-SILC, 2010) shows that there are significant proportions of households in which the partners' responses are not coherent, as well as of "full pooling" households in which at least one partner reports that they keep some of their own income from the pool (Ponthieux, 2013).

relationship, and hence do not provide accurate insights into the partners' living standards or control over money.

Another issue is whether the pooled system actually corresponds to the income pooling assumption. The few studies that have addressed this question seem inconclusive: Bonke and Uldall-Poulsen (2007) found a high correlation between the "pooled system" and an experiment-type measure based on answers to a question asking whether the respondents would spend more or less on themselves if their own and their partner's incomes were changed in opposite directions; partners using a pooled system were more likely not to change their personal spending. This result suggests that the two approaches correspond to the same notion. But Bonke (2013), using the same data set (the Danish expenditure survey) and questions allowing the household's consumption over a large set of goods to be assigned, finds that the structure of the household's expenditure seems to be correlated with the income distribution, even in households having reported a pooled system. In that case, the pooled system would not correspond to the assumption of income pooling.

12.2.5 Back to Measurement and the Pending Question of Intrahousehold Distribution of Income: Where We Stand, the Consequences, and Gendering the Issue

Concern about the distribution of income within the household is not new; in 1952, an article on "The distribution of income within the family" (Young, 1952) started as follows: "It is painfully obvious to the student of social policy that growing knowledge about the distribution of the national income between families has not so far been matched by a growth in knowledge about the distribution of the family income between its members. In place of knowledge, the assumption has often been made, though not stated, that the family . . . can still be treated as a unit for the purposes of spending. It has been taken for granted that some members of a family cannot be rich while others are poor. . . . To replace assumption by information is no small venture" (p. 305). More than 60 years later, it is painfully obvious that statistics on income and economic well-being remain based on the unitary version of the household,[22] as if stuck in a

[22] Nonmonetary indicators of poverty are not examined in this chapter, but it is worth mentioning that the measurement of material deprivation is, except in occasional surveys, also based on information collected from only one respondent, most often about "the household" in general or (rarely) about that respondent and their partner or children. With such data, the comparison of individuals within households is, of course, not possible (see, e.g., Adelman et al., 2002, who compare men and women living in different households). Using an Irish survey from 1999 allowing for intrahousehold comparisons, Cantillon (2013) founds that women tend to be relatively more deprived than their partners, especially concerning some specific items (for instance, going without a meal), and more significantly when the usual set of items used to measure deprivation is extended to leisure activities or personal spending, especially in couples with children. The combination of relatively low proportions of couples experiencing privation and the sensitivity of the responses to the conditions of the interview (whether the other partner is present) do not allow conclusions about clear patterns to be made—a point made in a previous study by Cantillon and Nolan (1998) using an earlier survey.

conceptualization now widely considered inappropriate. The gap between research and statistical practice has simply increased over time.

Statisticians are not unaware of the problem when deriving indicators of inequality between individuals from variables at household level. Although it is acknowledged that, in principle, the unit of interest is the individual, the argument that because incomes are shared the pertinent unit of collection is the household still holds[23] (Canberra Group, 2011, p. 24) or that, in doubt, it is preferable to assume that incomes are shared rather than the contrary (Förster and Mira d'Ercole, 2009, p. 7). So, at present, how to get a better idea of inequality in living standards within households remains an open issue. Taking a step forward in the conception of income data would entail a shift from household-level data to data systematically providing individual-level information on income components received or transfers made by and to individuals. There would, of course, be a cost in terms of data collection, and this would not in itself solve the question of the actual extent of sharing within households; but being able to use information disaggregated as much as possible would give more flexibility to testing various assumptions about income pooling and sharing and would allow escape from a situation in which the measurement of individual economic well-being and the analysis of inequality between individuals is conditioned in the first place by the assumption that individuals are equal within the household in which they are observed.

12.2.5.1 "How Serious Is the Neglect of Intrahousehold Inequality?"[24]

One major difficulty in assessing the potential consequences of ignoring inequality within households is, ironically, the lack of data on individual income. It is nevertheless possible to illustrate what the standard approach ignores when attributing an equal standard of living to each member of a household. We draw here on the framework proposed by Jenkins (1991, pp. 476–477). For a couple-household, a simplified formulation of the living standard (Eq-Y), taken as the household budget constraint, defines it as the sum of each partner's earnings plus the couple's net nonlabor incomes (NL) divided by the number of equivalent adults (n_{eq}) to account for the economies of scale[25]:

[23] It is interesting to note that the issue finally raised by the Canberra Group (2001) is not that of the assumption of income pooling and equal sharing within the household, but that of defining the pertinent income unit: "Then the relevant unit is either a person or group of related persons, within a household, whose command over income is shared" (p. 38), a condition that is taken for granted rather than checked.

[24] This title is borrowed from Haddad and Kanbur (1990). Using a survey of inequality in nutritional status in the Philippines, they found that neglecting intrahousehold inequality resulted in a significant underestimation of the level of inequality but did not drastically change the conclusions about the patterns of inequality—except between men and women.

[25] To simplify further, housework time is not taken into account.

$$Eq\text{-}Y = \frac{\left(w_f LM_f + w_m LM_m + NL\right)}{n_{eq}}, \tag{12.1}$$

where w is the earnings rate, LM the time in the labor market, and f and m identify females and males, respectively.

The standard approach assumes that $Eq\text{-}Y_f = Eq\text{-}Y_m$. This implies that any difference between $w_f LM_f$ and $w_m LM_m$ is counterbalanced by implicit transfers between the partners and that NL is received jointly or equally shared, and it assumes that the economies of scale are distributed equally between the partners. To take into account other neglected possibilities, the couple's total income, including the economies of scale (assumed separable), can be rewritten as

$$\begin{aligned} Eq\text{-}Y = & \left[a_1\left(w_f LM_f\right) + b_1(w_m LM_m) + a_2\left(NL_f\right) + b_2(NL_m)\right] \\ & + \left[\frac{(1-a_1)w_f LM_f}{n_{eq}} + \frac{(1-b_1)w_m LM_m}{n_{eq}} + \frac{(1-a_2)NL_f}{n_{eq}} + \frac{(1-b_2)NL_m}{n_{eq}}\right], \end{aligned}$$

where the first part of the right-hand side corresponds to the amount of incomes not pooled, and the a_is and b_is represent the share of income each partner holds back from the pool.

Compared with the standard approach, which assumes that any a_i, $b_i = 0$, and where any difference between w_f and w_m or LM_f and LM_m does not count since the incomes are fully pooled, leading to $Eq\text{-}Y_f = Eq\text{-}Y_m$, the second expression allows for a difference between $Eq\text{-}Y_f$ and $Eq\text{-}Y_m$ by an amount equal to $[b_1(w_m LM_m) - a_1(w_f LM_f)] + [b_2(NL_m) - a_2(NL_f)]$.

This amount illustrates the potential ignored inequality within a couple-household; it involves the degree of inequality between the partners' earnings, whether nonlabor incomes are received jointly or separately (which, in the case of capital income, raises the issue of asset ownership within multiperson households), and, of course, the share of his or her incomes each partner holds back from the pool. Whether this share is large or small remains unknown. However, the "pooling question" of the European survey "Intra-household allocation of resources" (EU Statistics on Income and Living Conditions [SILC] ad hoc module 2010) provides an order of magnitude of the proportions of individuals and households for which this share is not 0: over 21 European countries, the mean percentage of adults living in multiperson households who reported holding back at least some of their income from the pool was about 47%. At a household level and including one-person households (for which the assumption of full pooling is always true), this yields 38% of all the households in which at least one household member does not pool all his or her income (Ponthieux, 2013). Even though caution is necessary when interpreting the "pooling question" (see Section 12.2.4.2), this suggests a rather high risk of assuming full pooling when it could be unjustified. The next issue is that of the impact this may have on the measurement of economic inequality between men and women.

In the 1990s, several studies explored how men's and women's poverty rates would differ from the conventional measures under assumptions other than intrahousehold equality (Borooah and McKee, 1993; Davies and Joshi, 1994; Findlay and Wright, 1996; Fritzell, 1999; Phipps and Burton, 1995). The most general principle consists of applying an "unequal" distribution of income instead of the equal distribution assumed by the standard assumption; the "individual" income is then computed as if only the household nonlabor income was shared, using various assumptions about how it is distributed.[26] Whatever the year, country, or methodology, all these studies find that departing from the standard assumptions results in increased gender gaps in poverty (Table 12.2). In the case of married men and women, income poverty rates are dramatically higher for women and significantly reduced for men, instead of the equal poverty rates obtained with the conventional approach; when all households are included, the changes are less pronounced, mitigated by the relative share of one-person households for which the assumption has no effect.

Sutherland (1997) computed individual incomes assuming that all income is retained by the person who receives it, that the collective components of income are received by the head of household, and that family benefits are received by mothers—to whom Sutherland also allocates the responsibility for children. She shows that women are disproportionately represented in the bottom quantiles of the distribution of individual incomes, contrary to broadly equal shares of men and women in each quantile of the distribution of household income. A more recent study by Meulders and O'Dorchai (2010) suggested that the gap between conventional measures based on household incomes or "individual" incomes remains substantial. Using harmonized European data (EU-SILC) from 2006 for nine countries, they computed individualized incomes as the sum of the incomes received individually by adult men and women and, for those living in multiperson households, an equal share of the income components available only at the household level.[27] Then they compute a threshold at 60% of the median individual

[26] This is quite different from the approach developed in the literature on the "feminization of poverty," which focuses on households headed by single men and women, the argument being that couple households cannot, by definition, contribute to a gender gap in poverty. A limitation of these approaches is that they do not allow the effects of gender and the effects of the household composition to be sorted out (see Bennett and Daly, 2014, especially Section 12.5). The main results highlight that single mothers and single elderly women face higher-than-average poverty risks. Comparative work shows significant cross-country differences related to women's participation in employment and social policies (Barcena-Martin and Moro-Egido, 2013; Brady and Kall, 2008; Gornick and Jäntti, 2010; Kim and Choi, 2013; Wiepking and Maas, 2005).

[27] The income components that the data do not provide at individual level are capital income, net inter-household transfers, family benefits, taxes, and social contributions. Dividing equally the social contributions (which are based on labor incomes) is debatable, but the data do provide only an aggregate of taxes and social contributions.

Table 12.2 Sharing assumptions and poverty risks by gender

Sharing assumption **Authors, country, year, population**	**Equal sharing of household income**		**Other assumption on sharing**	
	Men	**Women**	**Men**	**Women**
Women get 30% of the couple's market income Borooah and McKee (1993), United Kingdom, 1985, married couples % below 2/3 mean equivalent income	33	33	14	66
Each adult keeps her/his own income[a] Phipps and Burton (1995), Canada, 1986, married couples % below 50% median equivalent disposable income	10.5	10.5	4.5	28
Each adult keeps her or his own income[a] Davies and Joshi (1994), United Kingdom, 1986, married couples % below 20th percentile equivalent disposable income	15	15	11	52
Women get 20% less than their equivalent income[a] Findlay and Wright (1996), United States, 1985; Italy, 1986; all adults % below 50% median equivalent disposable income				
Italy	17.5	16.8	15.4	27.1
United States	17.0	22.6	15.9	30.3
No sharing[a] Fritzell (1999), Sweden, 1991, adults aged under 65 % below 50% median equivalent disposable income	4	3.9	4.5	9.6

[a]The authors also present other assumptions or computations for other household demographic compositions than those reported in this table.

income and measure the risk individuals would face of falling below this threshold (called "financial dependency") if they lived only on their own financial resources.[28] The comparison of the two approaches shows that the conventional methodology has much more influence on women than men: over the nine countries, men's ratio of "financial dependency rate" to poverty rate ranges from 0.7 to 1.4 and women's changes from 1.7 to 3.3.

[28] This is not completely true because "individual" incomes are computed including an equal division of common income components, which are likely to be at least partly related to the current household composition (e.g., state transfers, taxes) and then change when the household composition changes. It is also perhaps not realistic to divide capital income equally because it is not necessarily common income. Another limitation is that, unless assuming that labor market status is totally independent of the individual's position in his or her household, the individual income is also likely to differ.

These results seem convincing enough to answer the question in the heading above: yes, neglecting intrahousehold inequality can be serious.[29] But while they show clearly that assumptions matter, it is more difficult to draw conclusions in terms of individual living standards; the main conclusion is that the living standard is a household-level notion and that household-level information fails to inform about individual situations within households.

12.2.5.2 The Standard Assumptions and the Assessment of Gender Economic Inequality

As exemplified above, by construction, the standard assumption of equality within households puts serious limits on the assessment of the extent of gender inequality. The crux of the problem is that the household dimension comes up entangled with individual characteristics, not only making an overall assessment of gender economic inequality virtually impossible but also obscuring the causal factors of gender inequality behind the household dimension. This is not to say that the household dimension has no effect on men's and women's economic outcomes—it certainly has strong effects, as will be obvious in the next sections—but the conventional measure of economic well-being conflates individuals and households, making individual situations intractable. The analytical challenge is to disentangle individual outcomes from household outcomes to understand how the "household dimension" both contributes to (because of decisions made in the household context) and obscures (because of assumptions about the absence of inequality within the household) gender inequality. This understanding may, in turn, have various implications in the perspective of public policies aimed at reducing inequalities.

The comparison of individual labor market status and outcomes and the risk of poverty is especially illuminating: in the working-age population, women, despite their lower participation in the labor market (see Section 12.3) and their disproportionate presence among low-wage workers when they do work (see, e.g., Grimshaw, 2011), do not face dramatically higher poverty rates than men: the gap (women − men) ranges from 2.2% to 2.9% points in English-speaking countries, between 0.9 and 2.1 in continental and southern European countries, and is negative in Nordic countries (figures are from Table 1 in Görnick and Jantii, 2010[30]). Moreover, Görnick and Jantii show that, among individuals of working age with low or no earnings, women are less likely to be poor than men. In a study including 22 EU countries, Maitre et al. (2012) found the same pattern among low-paid workers:

[29] The standard approach is also biased in a different way because the equivalent disposable income is measured directly for individuals who live alone and under strong assumptions for the others. These strong assumptions can themselves result in statistical artifacts, as can easily be illustrated by comparing the situation of two partners before and after separation (see Chiappori and Meghir's introduction to Chapter 16, in this volume).

[30] Their estimates are based on data from the early 2000s, but the order of magnitude remains comparable (for European countries) with most recent statistics from Eurostat (available from: http://epp.eurostat.ec.europa.eu/portal/page/portal/income_social_inclusion_living_conditions/data/main_tables).

men's poverty risk is notably higher than women's. A similar "gender paradox" is underlined in empirical studies of the working poor (Peña-Casas and Ghailani, 2011; Ponthieux, 2010). This, of course, has a lot to do with the fact that in the working-age population, a large proportion of men and women live in couple households, a configuration in which there is no inequality by construction. The measure of women's economic well-being then seems to be much less related to the outcome of their individual economic activity than men's. However, this is precisely because their earnings are, on average, lower than men's. Because of this inequality and the large share of adults who do not live alone, the influence of the equal sharing assumption is the greatest on the estimates of women's poverty rates.

What does this imply in terms of public policies? Many social transfers target the households (families, fiscal units) and assess their efficiency on the basis of outcomes measured at the household level. The implicit assumption is that meeting the needs of individuals is achieved by meeting the needs of households—another formulation of the unitary model. It is the same when individual benefits are conditioned by household resources, implicitly assuming that there cannot be needy individuals in nonneedy households. How this may affect the economic well-being of the individuals within households is, by construction, difficult to assess precisely because conventional measures of economic well-being and policy targets are derived from household-level information. While no statistical information allows for an assessment of the incidence of individual poverty in nonpoor, multiperson households, there are reasons to believe that neglecting the intrahousehold distribution of resources can contribute to the persistence of gender inequality. Policies that link what an individual is entitled to with the actions or resources of another member of his or her household can reinforce the imbalance of resources between men and women and women's economic dependence.

However, even accepting that intrahousehold transfers are such that equal living standards are achieved despite unequal personal resources, and that, contrary to what empirical results suggest, the imbalance of resources within households is not an issue, the equilibrium can be fragile. The trend of decreasing marriage in favor of cohabitation and increasing divorce or separation rates results in an increase in the number and types of households an individual may belong to over his or her life cycle. This suggests that resorting to household outcomes to assess individual situations may become increasingly irrelevant. From the perspective of measuring and analyzing gender inequality, this calls for a change of framework; as long as women's labor market outcomes are less favorable than men's, as is still the case, the standard approach, bypassing the issue of intrahousehold inequality, conceals gender asymmetries in the relation between the market and the household and inequality in terms of autonomy, risks, and opportunities.

12.3. THE GENDER WAGE GAP

Women's increasing participation in the labor market was a major trend in the second half of the twentieth century. At the same time, women's educational levels caught up with those of men—even overtook them, on average, in nearly all advanced countries—and women gained access to a greater number of occupations, favored by technological change. Since the 1970s, the effects of what Goldin (2006) describes as a "quiet revolution" have changed not only women's expectations about their careers and family lives but also public attitudes toward women's social role. Nowadays, in most industrialized countries, the participation of women, and especially mothers, in paid employment has become almost a norm; dual-earner families outnumber "male breadwinner" families; gender equality is an explicit policy goal, and gender discrimination is prohibited by law in many countries. But women's outcomes in the labor market still remain significantly lower than men's.

This section presents an overview of the analysis of gender inequality in wages, focusing particularly on the recent search for new explanations of its persistence. Over recent decades, this gap has narrowed in many OECD countries, but the trend of convergence now seems to be very slow, if not stationary (Blau and Kahn, 2006a,b). Since the early 1970s, the fact that men earn higher wages than women on average has usually been analyzed as the result of two main factors: a gender gap in productivity (due to women's lower levels of human capital and/or effort at work) and gender discrimination (cf. the survey by Altonji and Blank, 1999). With women's human capital (education and experience) catching-up men's, these variables become ever less relevant to explaining the gender wage gap. The stylized fact today is that gender wage inequality results from gender differences in occupations and from the lack of women at the top. This raises the question of why women still do not have similar careers to men. A vast literature competes in proposing explanations that are generally nonexclusive, complex, and of a scope far beyond standard human capital models. A growing strand of this literature incorporates psychological factors and social norms, shedding new light on the relationships between individual behavior and labor market outcomes (see a survey by Bertrand, 2010). But this does not rule out the central role played in the gender wage gap by the unequal share of unpaid work. Children are still at the heart of gender inequality on the labor market, and family constraints are largely responsible for career interruptions, part-time jobs, reduced working hours, and slow wage progressions. Before reviewing these developments, we start with a brief overview of the main trends and cross-country differences since the 1990s.

12.3.1 Main Trends and Cross-Country Differences in Gender Labor Market Outcomes

Wages are among the most available data on income by gender, but comparing wages (and the gender wage gap) over time and between countries presents a number of difficulties. Ideally, comparisons of the gender wage gap should be based on data covering all employees, sectors, and industries and a consistent concept of wage. Such data do not exist. National sources do not necessarily provide directly comparable information because of the variety of possible statistical sources (self-reported or administrative data), differences in scope (e.g., covering some or all sectors and industries, including part-time workers or not), differences in the concept of wage (especially whether it includes overtime, bonuses, etc.). Indicators of the wage gap can be calculated on the basis of hourly, weekly, monthly, or yearly wages, which of course do not yield the same result; they can cover all the employees or (and most frequently) only full-time workers; they can be calculated as a differential of the mean or median wages. It is therefore not surprising that the size of the gender wage gap can vary even within the same country and over a single year, making comparing gender wage gaps between countries even more tricky (on the importance of comparable data for the measurement of the wage dispersion in international comparisons, see Salverda and Checchi, 2014, Chapter 18, in this volume).

12.3.1.1 Narrowing, But at Different Speeds Across Countries, and Not Closing Completely

The overview of trends and cross-country differences presented below does not escape these limitations. We use the information available from the OECD. The information from the OECD has the advantage of scope and time coverage (1975 through 2011, although not for all countries); one drawback is that information is based on national definitions; another drawback is that it only covers full-time (or, in a few cases, full-time equivalent) employment, hence making it impossible to assess the impact of part-time work on the gender wage gap. Graphs are the easiest way to "see" trends over time, so Figure 12.2 displays the average gender pay gap over half-decades since 1975 for a sample of OECD countries, using a 4-year moving average to smooth the changes.

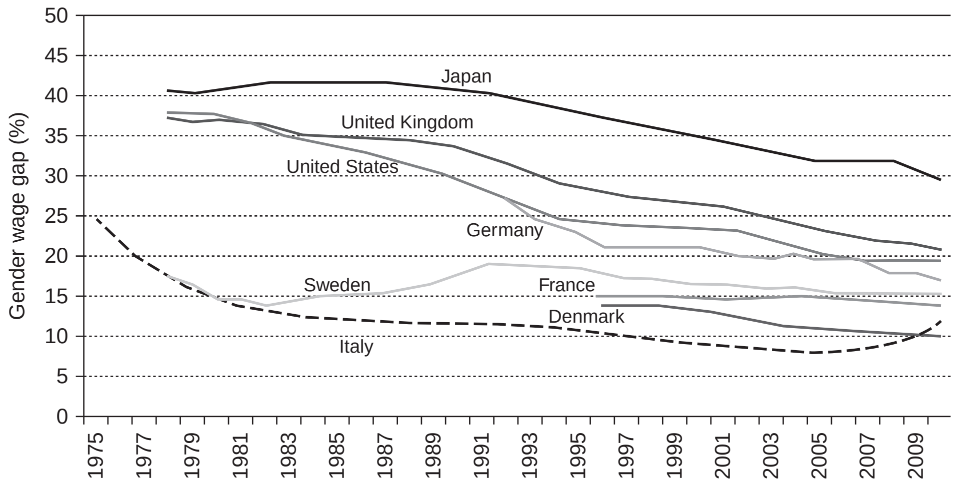


Figure 12.2 The gender wage gap in selected Organisation of Economic Co-operation and Development (OECD) countries from 1975 through 2007 (4-year moving average considering full-time workers). The wage gap is measured as the mean men's wage minus the mean women's wage divided by the mean men's wage. The gap is measured for various concepts of wage (annual, monthly, hourly), depending on the country. *Source: OECD, employment database 2012, national definitions and sources. Available from: http://www.oecd.org/employment/emp/onlineoecdemploymentdatabase.htm#earndisp.*

The general view of a wide gender pay gap narrowing fast and then more slowly is clearly verified for the United States, whereas the pattern is less uniform in other OECD countries. The United States experienced the most dramatic narrowing of the gender wage gap during the 1980s. This was particularly remarkable because it occurred during a period when the level of labor market inequality was rising sharply in the United States. This movement was called "swimming upstream" in a famous paper by Blau and Kahn (1997): women's improving qualifications, especially their experience and occupations, made up for the rising inequality. In the OECD countries, the gender wage gap has been decreasing more slowly since the late 1990s (except in the United Kingdom and Japan, where the narrowing has continued at the same pace) or stagnating, and even increasing in Italy.

Historically, the main factors behind the general narrowing of the gender wage gap have been the increases in women's educational levels (Goldin et al., 2006) and in labor market experience together with the associated returns (O'Neil and Polachek, 1993). Since the 1970s, the catching-up of women's educational levels and a reversal of the gender education gap have been observed in most OECD countries. Women now outrank men in 29 of the 32 OECD countries; in 2009, about 60% of all graduates with a university degree were women (OECD, 2012).[31] However, the specialty of degree remains highly gendered: only 26% of graduates in engineering, manufacturing, and construction are women, compared with more than 75% in health and welfare (OECD, 2011a). This has very important implications for occupational segregation (see below).

The increase in women's participation in and attachment to the labor market, and the corresponding increase in their work experience, is believed to result from changes in the demand for labor and in women's labor supply. Labor demand changes were driven mostly by major technological changes such as computerization[32] and the rise of the service sector. On the supply side, two sorts of technological change may also have contributed to women's increased attachment to the labor market by modifying their constraints: progress in domestic technology, reducing the time required for domestic chores (a debated factor; see Section 12.6), and the contraceptive pill, of which Goldin and Katz (2002) underline the central role in allowing women to postpone maternity, spend more time in education and at work, and "form their identities before marriage and family" (Goldin, 2006, p. 14), and which was a true cultural change. However, the impact of these changes (which occurred with varying delays and intensities outside the United States) is a sort of one-shot impact and cannot explain the lasting trend of convergence between women's and men's participation rates over the 1990s and

[31] OECD, http://www.oecd.org/edu/skills-beyond-school/49986459.pdf.

[32] Weinberg (2000) finds that computerization may account for more than half the increase in demand for female workers in the United States.

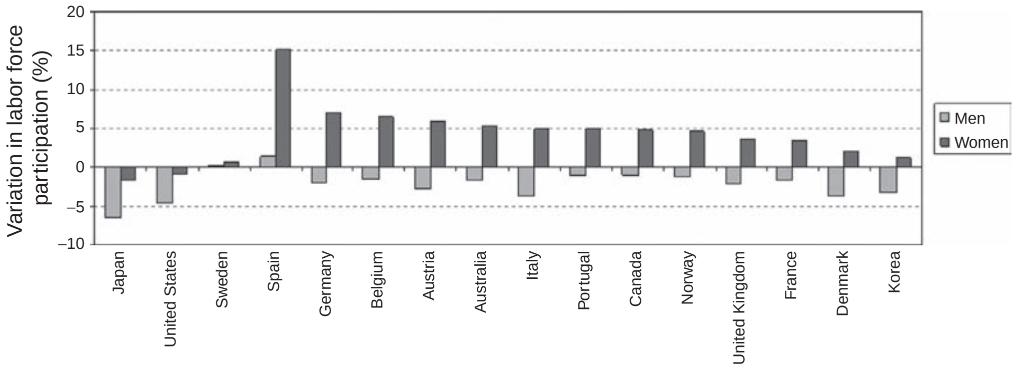


Figure 12.3 Variations in men's and women's labor force participation, 1990–2011. *Source: Organisation of Economic Co-operation and Development, employment database 2012, http://www.oecd.org/gender/data/harmonisedindicators.htm.*

2000s.[33] In most countries, this trend stems from the continuing increase in women's participation in the labor force, combined with a slight decrease in men's participation (Figure 12.3). The exceptions are Japan and the United States, where both participation rates have decreased (but men's more than women's); Sweden, which has hardly changed; and Spain, where men's participation increased, albeit far less than women's.

12.3.1.2 Gender Labor Market Outcomes: An Overview for OECD Countries

At the end of the 2010s, the average gender wage gap in the OECD countries is about 16%, but there are wide differences across countries: it is more than 25% in 3 countries, between 20% and 25% in 3 countries, between 15% and 20% in 9 countries, between 10% and 15% in 10 countries, and less than 10% in the remaining 7 countries (Table 12.3).

Several factors can help to explain these differences. The most important of these factors is the degree of inequality in the national wage structure. Other possible explanations, but at a lesser level, are gender differences in the employment rate, part-time work and the unemployment rate. We will briefly review differences across countries for these indicators.

12.3.1.2.1 The Wage Structure

Because women tend to be concentrated at the bottom of the wage distribution, a more compressed male wage structure reduces the gender pay gap, as shown by Blau and Kahn (1992, 2003): the average wage penalty resulting from women's unfavorable ranking in the male wage distribution is mechanically smaller in countries where the wage distribution is less widespread (Blau and Kahn, 1996). One striking example of this mechanism is given by Blau (2012, pp. 140–141), who compares Sweden and the United States. Gender earnings ratios (adjusted for hours) are 77.3% and 65.4%, respectively, and women's

[33] This continuous increase of women's participation rate is probably mostly driven by the expansion of work–family conciliation policies in most OECD countries (Blau and Kahn, 2013).

Table 12.3 Organisation of Economic Co-operation and Development countries grouped by size of the gender wage gap (among full-time workers),[a] 2010
Gender wage gap

<10%	10% to <15%	15% to <20%	20% to <25%	>25%
Denmark	Slovak Republic	United Kingdom	Israel	Korea
Norway	Sweden	Austria	The Netherlands	Estonia
Belgium	Iceland	Finland	Turkey	Japan
New Zealand	France	United States		
Luxembourg	Australia	Canada		
Hungary	Portugal	Switzerland		
Poland	Greece	Germany		
Spain	Slovenia	Chile (2011)		
	Ireland	Czech Republic		
	Italy			

[a]Countries are ordered in each group from the highest to the lowest gap.
Source: Online Organisation of Economic Co-operation and Development Employment Database. http://www.oecd.org/els/emp/onlineoecdemploymentdatabase.htm.

mean ranking in men's wage distributions[34] is 28.2 in Sweden and 32.3 in the United States; the less favorable position of women in the wage hierarchy in Sweden nevertheless results in a lower pay gap than in the United States because the wage distribution is more compressed in Sweden.[35] The wage structure compression itself depends essentially on differences in wage-setting institutions: highly centralized, unionized wage settings reduce the wage dispersion and have a positive effect on low-paid workers, who are predominantly women; similarly, the existence of a minimum wage raises women's relative pay by limiting its lowest possible level (on labor market institutions and wage dispersion, see Salverda and Checchi, 2014, Chapter 18, in this volume).

12.3.1.2.2 The Gender Gap in Employment

Despite the trend of convergence, women's employment rates generally remain significantly lower than men's (Figure 12.4), and it has been found that the gender wage gap could be negatively correlated with the gender gap in employment. For instance, the employment gap is much smaller in the United Kingdom than in Italy, and the wage gap is much smaller in Italy than in the United Kingdom (cf. Table 12.3). This is caused by effects of selection into employment, as shown by Olivetti and Petrongolo (2008): the

[34] Each woman is assigned a rank in the male wage distribution according to her wage. The mean female percentile in the male distribution is the mean of these rankings. If women's wage distribution were exactly similar to men's, the mean would be equal to 50. The lower the women's mean ranking, the more unfavorable their position in the wage hierarchy.

[35] Women's ranking in men's wage distribution is a useful indicator of gender differences in wage hierarchy and complements the average gender wage gap, which gives the monetary effect of the ranking inequality. Unfortunately, this indicator is rarely available in published statistics.

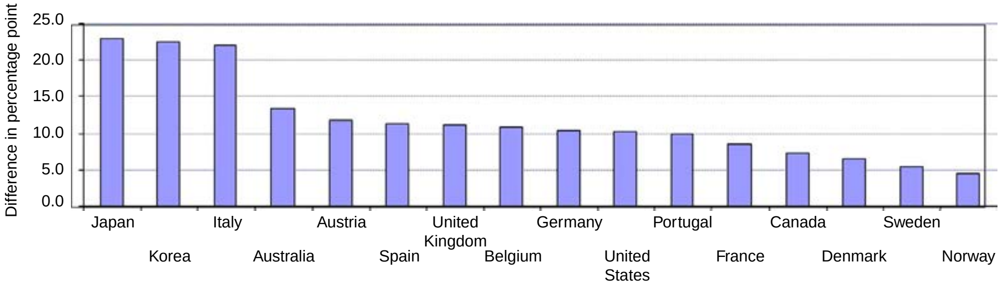


Figure 12.4 Gender gaps in employment in selected Organisation of Economic Co-operation and Development (OECD) countries, 2011. *Source: OECD Employment Outlook (2012).*

fewer women employed, the more they are positively selected and the higher their relative wage. In a study of the impact of reunification on former East Germany, Hunt (2002) reported a similar effect. She found that in the 4 years following reunification, the gender wage gap dropped by 10% points. But she also found that, because of a decrease in the demand for low-skilled workers, the employment rate of women had fallen by 6% more than men's and that this differential in exits from employment explained half the relative wage gain of women.

However, the relation between women's participation in employment and the gender wage gap is not uniform, as shown by the comparison between Italy and Japan, two countries that are very similar in their high gender employment gaps (the highest of the countries displayed in Figure 12.4) but are far from similar in their gender wage gaps (see Table 12.3). While employed women in Italy are becomingly increasingly skilled compared with "random" employed men, however, the gender gap in education remains pronounced in Japan, and employed women are mostly concentrated in part-time or short-term jobs, with men securing the best and best-paid jobs (see Chang and England, 2011, on the heterogeneity with regard to gender inequality within Asian industrialized countries).

12.3.1.2.3 Part-Time Work

Part-time work is a typical feature of women's employment. The percentage of part-time work among men is growing slightly, but at the end of the 2000s, it remains, on average, well below 10% in most OECD countries (with the noticeable exception of the Netherlands), while it represents anything from 15% to about 40% among women and up to 60% in the Netherlands. Comparing total employment or full-time employment may lead to very different views of cross-country differences. Figure 12.5 shows the gender gaps in employment in the OECD countries (OECD, 2013) using two measures: the gap in total employment and a gap measured in full-time equivalent employment (part time is defined as less than 30 weekly hours). The contrast between the two indicators reflects the difference in the share of part-time employment between women and men, which can be very large, as for instance in the Netherlands, Germany, and Switzerland and, to a lesser degree, in Austria and the

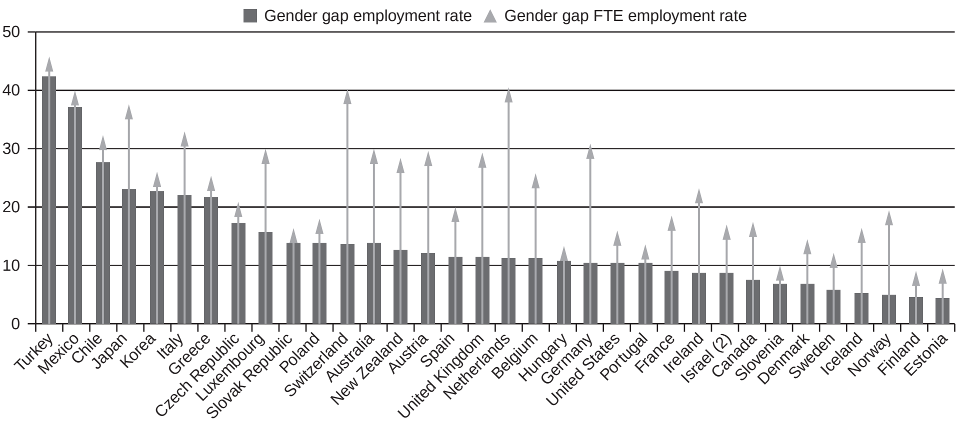


Figure 12.5 The gender gap in employment rate and in full-time equivalent (FTE) employment rate, 2011. *Source: OECD Family database www.oecd.org/social/family/database (2) The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities http://dx.doi.org/10.1787/888932315602.*

United Kingdom. Part-time employment, which contributes to women's participation in the labor market—especially by making it easier to balance work and family constraints (see Jaumotte, 2003)—may have adverse effects on women's careers and wages (see below). Unfortunately, as mentioned earlier, indicators of the gender wage gap, including part-time employment, are not available at the OECD level.

12.3.1.2.4 Gender Gaps in Unemployment

In most OECD countries, the unemployment gap has virtually disappeared as a result of women's increased labor force attachment (Albanesi and Sahin, 2013). In the United States there has been no difference between men's and women's unemployment rates since the 1980s —except during recessions, when men's unemployment rates exceed women's (but the recovery is also faster for men). In the European Union, unemployment rates have converged since 2000, and the gap can be said to be mostly residual (Eurostat), with the exception of southern countries (Greece, Italy), where women's relatively low attachment to the labor market makes them more at risk of unemployment.[36] Azmat et al. (2006) also found a prominent effect of women's participation rate on the gender gap in unemployment and explain cross-country differences by the interaction of gender differentials in labor market attachment and institutional factors—plus perhaps hiring. Institutions contributing to a more compressed wage distribution (minimum wage, unions) or to a reduction of labor turnover (firing costs and subsequent reduced

[36] However, we must bear in mind that the overall unemployment rate may influence the decision to participate in the labor market: expectations of a high risk of unemployment may discourage labor supply in the same way as other anticipated work interruptions. The resulting selection effect in labor supply may then reduce the gender unemployment gap.

hiring rates) as well as temporary contracts (making it difficult to stay in employment) increase the unemployment rate of low-skilled workers or workers with less attachment to the labor market. Women (whose outflow rates from employment are higher than those of men) are then more likely than men (more often employed in permanent jobs) to be affected by reductions in hiring.

To sum up: at the end of the first decade of the twenty-first century, the raw gender wage gap stands between 10% and 20% in most industrialized countries, but it has been at this level since the mid-1990s; men's and women's employment rates continue to converge, but women are still much more likely to work part time.

12.3.2 Gender Segregation on the Labor Market

Not so long ago, it was legal to offer women lower wages than men for the same job or to hire only men or women for certain jobs. In the United States, job advertisements could be published separately by sex until 1963, the year of the Equal Pay Act, which prohibited wage differentials based on gender. In Australia, the law allowed employers to pay women up to 25% less than men for the same work until 1969. It was the same in New Zealand, where the legal pay gap was set at 50% in 1903, lowered to 30% during World War II, and finally abolished in 1972. Equal pay for equal work is now the law in many countries, and the antidiscrimination legal framework has been reinforced in recent decades. In theory, pure wage discrimination is no longer possible.

The persistence of a gender wage gap between workers with similar productive characteristics but of different genders is therefore puzzling; it suggests that there is still a possible discrimination, but one different from pure wage discrimination (that is, unequal pay for equal work) and difficult to assess. Wage differentials between apparently comparable workers can result (and actually do result) essentially from the unequal distribution of men and women across occupations, firms, or sectors more than from differences in pay rates within occupations in a same firm (see Blau, 2012, specifically chapter 1, "Equal Pay in the Office"). In this section we start by briefly describing the tools of empirical research on the gender wage gap. Then we turn to analyses of the effect of gender segregation by firm, industry, and occupation. Finally, we present the important strand of literature devoted to vertical segregation and the debates around the "glass ceiling effect."

12.3.2.1 How to Analyze the Gender Wage Differential

Since the 1970s, numerous empirical studies of the gender wage gap have sought to quantify the contributions of various factors to pay differences between men and women or to the way they have changed over time. Fundamentally, this means distinguishing between the part of the gap that is due to differences in the observable productive characteristics of men and women and the part that is not.

The standard methodology used in this huge number of studies is decomposition of the gender wage gap, initiated by the seminal papers by Oaxaca (1973) and Blinder (1973). The basic principle is to rewrite the gender wage gap as the sum of two differences: a difference in the "quantities" of men's and women's productive characteristics (i.e., human capital variables) and a difference in the prices of these characteristics as measured by the difference in their estimated returns. The decomposition method has been developed and extended to explain changes over time in the unexplained wage gap (Juhn et al., 1991, 1993); to integrate quantile analysis (Albrecht et al., 2003; Machado and Mata, 2005); and to treat dichotomous outcomes (Fairlie, 2005), nonlinear models, censored outcomes (Bauer and Sinning, 2008), and nonparametric setups (Ñopo, 2008a,b). The case of distributional parameters other than the mean has been incorporated into this framework, and different methods have been proposed to study the entire conditional distribution (Chernozhukov et al., 2013; DiNardo et al., 1996; Firpo et al., 2009; Fortin and Lemieux, 2000). Decomposition methods still form a very active field of research in econometrics (see the survey by Fortin et al., 2011).

Decomposition methods provide a useful tool to assess the contributions of various factors to the gender wage gap and to diagnose which of these factors are the most quantitatively important, but they do not say anything about the underlying economic mechanisms. For instance, the variable "occupation" is responsible for a large part of the gender wage gap, but this difference may result from employer's discrimination toward women or from gender differences in the choice of occupations; clarification of this point requires further analysis.

Applied economists face some practical problems when decomposing the gender wage gap. One classic difficulty is that the measurement error of some key variables may be more marked for women than for men. This is particularly likely for the measure of actual labor market experience, which is not always available in individual data sets and in this case is replaced by a measure of potential experience (current age minus school-leaving age). Because women are more likely than men to experience career breaks, their potential experience is overestimated, biasing downward the returns to experience and so biasing upward the unexplained part of the wage gap (cf. Regan and Oaxaca, 2009, who used the 1979 National Longitudinal Survey of Youth and the Panel Study of Income Dynamics). Another difficulty in choosing the specification of the wage equation is the risk of omitting variables that could be correlated with the included variables; this bias can work in either direction (Neumark and Korenman, 1992).

Workers, especially women, are generally not a random sample of the working-age population. Certain unobserved factors may determine whether an individual is employed, and they may be correlated with the individual's productive characteristics and the wage. Following Heckman (1976, 1979), sample selection bias is treated by introducing a correction term (the inverse Mills ratio), obtained from a probit model of the probability of being employed, into the wage equation (Neuman and Oaxaca, 1998).

This method itself introduces a practical difficulty: the need to find at least one explanatory variable correlated with labor market participation but not with wage (i.e., an exclusion variable; see Vella, 1998).[37]

Another standard issue is that of the "norm" (price) to be used in the valuation of the difference between productive characteristics. In the initial formulations of Oaxaca and Blinder, the coefficients estimated for men were used as the price norm, which was arbitrary (technically, they could equally well have used the estimated returns to women's endowments) but fairly logical given the sign of the wage gap. However, using men's or women's coefficients as prices does not result in the same estimates of the explained and unexplained parts of the gap. Neumark (1988) and Oaxaca and Ransom (1994) proposed a generalized method using the coefficients obtained by estimating the wage equation over the pooled sample of men and women as the norm (alternative weighting schemes have been proposed by Cotton, 1988 and Reimers, 1983). Fortin (2008) updated the Neumark (1988) or Oaxaca–Ransom approach (1994) by including gender intercept shifts in the pooled regression with an identification restriction, so that the nondiscriminatory wage structure is such that the advantage of men equals the disadvantage of women.[38]

Such decomposition techniques have been widely used on data from different countries, periods, and samples of workers. Given all the possible variations in the data collected and the methodology adopted, the results are difficult to compare, even within a single country. A meta-analysis by Weichselbaumer and Winter-Ebmer (2005) using 263 published papers covering 63 countries between the 1960s and the 1990s[39] offers a useful overview of these findings and their sensitivity to the methods used. Ultimately, data restrictions (how wages are defined, which sample is studied, the quality of independent variables) were found to be more important than the choice of econometric methods in the decomposition of the wage gap. They found that the overall decrease in the total gender wage gap (from 65% in the 1960s to 30% in the 1990s) was mostly explained by the improvement in women's productive characteristics (education, experience). The decrease in the explained part is not surprising; as mentioned above,

[37] Blundell et al. (2007) proposed an approach for dealing with nonrandom selection when there are no obvious identification strategies; they developed bounds based on restrictions derived from economic theory (one of their restrictions is that the probability of working is higher for those with higher wages) and used their methodology to analyze changes in gender wage inequality in the United Kingdom.

[38] Another well-known problem in the Oaxaca–Blinder decomposition is that the portion of the explained part of the gender wage gap assigned to a categorical variable depends on the omitted category. To solve this problem, Gardeazabal and Ugidos (2004) suggested including all categories in the wage regression and imposing coefficients on a set of indicator variables to sum to zero. The method proposed by Fortin (2008) also includes this constraint for each categorical variable.

[39] All these papers used the Oaxaca–Blinder type decomposition. Because US papers were overrepresented at the beginning of the period (65% of the papers based on data for the 1960s), the authors applied a weighting scheme to correct for this bias.

women's education levels are comparable to those of men in OECD countries, and because their increasing participation in the labor market has given them similar levels of work experience,[40] the gender wage gap is less able to be explained by gender differences in basic human capital variables. The remaining component is predominantly due to occupational segregations: women and men do not occupy the same jobs (horizontal segregation) or have the same wage careers (vertical segregation). In both cases, these differences are unfavorable for women, but the gender gap in promotion and access to top jobs is generally considered the main cause of gender pay inequality.

12.3.2.2 Gender Differences in Occupations and Sectors

Occupational gender segregation in the whole economy is observed in all OECD countries: less than 10 occupations[41] accounted for half of the total employment of women in 2009, compared with 12 or more for men (OECD, 2012). A similar pattern is observed for the United States but at a lower degree than in most industrialized countries (Anker, 1998). Dolado et al. (2003) used a common detailed classification (108 occupations) to compare horizontal segregation in the United States and Europe in 1999. The degree of occupational concentration is smaller in the United States than in Europe,[42] but in both cases women are concentrated in similar types of jobs (salespersons, domestic help, personal care, secretaries, and teachers).

12.3.2.2.1 The Multiple and Complex Causes of Gender Differences in Occupations

The first explanation of this general finding is persistent gender differences in the field of study; in short, few women choose science or technology, and this has direct consequences on occupational segregation[43] and, consequently, on wages. The choice of subject of study is, then, a pre–labor market decision that has significant long-term consequences on earnings. The effects of college major on wages have been studied for the United States (Black et al., 2008; Brown and Corcoran, 1997; Loury, 1997; Weinberger, 1998, 1999); these studies concluded that it has a significant effect on early career wages. Including the field of study in the wage equation considerably improves the explained part of the gender pay gap. Similar results were obtained by Machin and Puhani

[40] The remaining gender differences in work experience are mainly due to family constraints: women are still more likely than men to have a discontinuous work life because they are more responsible for caring for children (or dependent adults). Part of the gender wage gap is then related to the impact of children on their work life and their chances of reconciling work and family. We return to the specific question of career interruptions and their impact on wages in the later section on the family pay gap.

[41] This indicator depends on the choice of classification of occupations and the degree of detail; here the 3-digit ISCO88 classification of occupations has been used, distinguishing 111 occupations.

[42] Two-thirds of female workers are concentrated in 21 occupations in the United States and 18 in the European Union.

[43] Even the link between field of study and occupation is weaker for women than for men: those who choose science, technology, or mathematics are less likely to pursue a science career than men (OECD, 2012).

(2003) for Germany and the United Kingdom: the choice of degree explains a significant part (between 8% and 20%, according to the specifications) of the overall gender wage gap among graduates. This raises the question of why women are so reluctant to choose science, given the well-known consequences on their future earnings and their high performance at school. Traditionally, differences in the choice of study subjects and their consequences on chosen occupations have been explained by taking into account the outside options for women, i.e., family responsibilities. Polachek (1981) argued that women choose occupations (and the corresponding investment in education and training) in which the cost of career interruptions is low. Because women's work attachment is stronger nowadays and their careers are more often continuous, this kind of explanation is less convincing. Economic analyses are turning toward more psychological approaches and gender stereotypes to explain this conundrum (see Section 12.3.3). The choice of field of study may also depend on the environmental conditions of schooling and peer effects. In line with educational studies research, recent economic papers test the causal impact of the gender composition of classes or teachers in the choice of field of studies. This is done using random variations in teacher gender (Carrell et al., 2010) or in the gender composition of classes (Schneeweis and Zweimüller, 2012). Results tend to show that exposure to a female-dominated environment encourages girls toward male-dominated fields.

In addition to investment in education, occupational segregation results from both supply and demand factors. On the supply side, women may prefer a given occupation because of nonmonetary advantages (flexible work hours, job amenities), in line with Adam Smith's compensating differentials theory, in which case they trade nonmonetary advantages for lower wages. For instance, Filer (1985) and MacPherson and Hirsch (1995) support the explanation of gender differences in preferences for occupational characteristics and found that taking into account occupational characteristics reduces significantly the unexplained part of gender wage differentials. But, as in the case of educational choice, women may also choose a given occupation to comply with social norms and stereotypes.

On the demand side, discrimination against women (taste discrimination) or the employer's perception that women are, on average, less productive or reliable than men (statistical discrimination) may affect the gender composition of occupations. Hiring discrimination has been estimated by correspondence testing (Booth and Leigh, 2010; Duguet et al., 2005; Petit, 2007; Riach and Rich, 2002) on the French financial sector) or audit studies (Neumark, 1996 on waiters and waitresses[44]). Correspondence testing shows that women are less likely to be selected for an interview in high-status jobs and in traditional male occupations; the opposite is observed in female-dominated occupations (secretary, for instance). Weichselbaumer (2004) extended correspondence

[44] They found that women have a lower probability of being interviewed for a job in high-price restaurants.

testing to investigate whether gender stereotypes affect the hiring process. She included personal characteristics given in curriculum vitae (hobbies, photographs) and did not find that a masculine identity for a woman (indicated by a masculine look and hobbies, such as motorbiking) reduces unfavorable treatment in masculine occupations; she concluded that discrimination is largely responsible for sex segregation. In that case, the proportion of women in a given occupation may change if the hiring process is modified and a fairer procedure is adopted. A famous example of this mechanism is presented by Goldin and Rouse (2000), who examine the effect of the adoption of "blind" auditions behind a "screen" to recruit musicians to American symphony orchestras. The blind audition procedure clearly increased the proportion of women in orchestras, suggesting that hiring discrimination was responsible for the previous quasi-absence of women in orchestras.

12.3.2.2.2 A Slow Decrease in Occupational Segregation

Occupational segregation has changed over time. The increase in women's education and labor market participation since the 1970s has modified the quality and quantity of the female labor supply. It is also possible that hiring discrimination decreased with the strengthening of antidiscrimination policies. Conversely, the expansion of the service sector, partly driven by the marketization of housework (Ngai and Petrongolo, 2013), and technological changes biased in favor of highly skilled workers have affected the occupational structure of female employment and their tasks (Black and Spitz-Oener, 2010).[45] Occupational segregation in the United States tended to decline slowly in the 1960s (Blau and Hendricks, 1979), and this trend accelerated over the 1970s (Bianchi and Rytina, 1986). It has continued to decline, but at a much slower pace, since the mid-1990s (Blau et al., 1998) or even stagnated (Hegewisch et al., 2010). The study of occupational segregation by gender in the United States from 1970 to 2009 by Blau et al. (2013), with a consistent set of occupational categories, confirmed the slowing decline of occupational gender segregation. This study also confirmed that the reduction in occupational segregation is positively correlated with education. The largest decrease is observed among college graduates, whereas there is a very limited change among high school dropouts. This result is in line with the results of Black and Juhn (2000), who observed, again for the United States, a strong increase in the percentage of college-educated women in high-paying professional occupations (increasing from 8% in 1967 to 23.5% in 1997). Their analysis was that women had positively responded to the increase in demand by firms and increasingly chosen "career jobs," either by

[45] Using a West German database, Black and Spitz-Oener (2010) examined changes in tasks consecutive to technological progress and the spread of computing. They found that women experienced a strong decline in routine tasks (with a positive effect on wages), contrary to men.

increasing their labor market participation or by moving away from traditional "female" occupations (nursing, teaching).

Occupational segregation is related to workplace gender segregation in a complex interaction. Using matched employee–employer data for the United States, Hellerstein et al. (2007) found a decrease in workplace segregation by sex between 1990 and 2000, contrary to workplace segregation by race or ethnicity. Interestingly, changes in the occupational distribution of men and women were not the main reason for the decline in workplace gender segregation. The main driving force was the change of occupational structure within firms, with a growing share of mixed occupations and the decline of female- or male-dominated occupations. Moreover, the rise of the service sector (in which women represent a large majority of the workforce) has slowed the decline of horizontal segregation.

12.3.2.2.3 Do Female-Dominated Occupations Systematically Pay Less?

What are the consequences of this occupational segregation for the gender wage gap? It is generally considered that female-dominated occupations pay less, following Bergmann's (1974) "overcrowding" model: the discriminatory behavior of some employers creates an excess supply of women, which depresses their wages for a comparable occupation. A wide empirical literature on the US case verifies this prediction: controlling for measured characteristics of workers, predominantly female occupations pay less than predominantly male occupations. Consequently, occupational gender segregation accounts for a part of the gender wage gap (Bayard et al., 2003; Groshen, 1991; Killingsworth, 1990[46]). For the United States, Groshen (1991) concluded that the importance of occupational segregation in the gender wage gap is an argument in favor of comparable worth policy (equal pay for work of equal value) as a powerful tool to reduce the gender wage gap. Amuedo-Dorantes and De la Rica (2006) also found a notable effect of female segregation into low-paying jobs on the gender wage gap in Spain. This effect can be amplified by gender segregation in firms (particularly small firms; see Carrington and Troske, 1995) and the corresponding wage penalty. Because women are more likely to be employed in low-wage workplaces, the unexplained gender wage gap is reduced when a workplace fixed effect is included[47] (Drolet and Mumford, 2012).

But the penalty attached to female-dominated occupations is not observed everywhere, or only to a lesser extent. In their comparative study of the United States and

[46] Bayard et al. (2003) found also a large within-occupation/within-firm gender wage gap; this results in part from using broader occupational categories compared with other studies.

[47] Note that the workplace-specific return may be different for men and women. Meng and Meurs (2004) for Australia and France and Drolet and Mumford (2012) for the United Kingdom and Canada found a higher workplace return for women than for men. In other words, firms' policies tend to narrow the gender wage gap. This effect is particularly strong for the private sector in Canada, which may be due to Canada's pay equity legislation.

Canada, Baker and Fortin (1999) found a small and not statistically significant effect of occupational gender composition on Canadian women's wages; they attribute this result to a higher unionization rate in Canada than in the United States. Jurajda and Harmgart (2007) studied the interesting case of former West and East Germany and did not find a wage penalty on female jobs in West Germany; in East Germany predominantly female occupations offered higher wages.[48] Female-dominated occupations often are associated with jobs providing care for others, such as childcare, nursing, teaching, or social work. Again, the wage penalty attached to these jobs is not uniform, nor is it observed in all countries. Barron and West (2013) examined the gender wage penalty in caring occupations in Britain and found that this penalty is located in occupations requiring lower levels of educational qualification (childcare workers, nursing assistants, auxiliaries) and not in highly qualified occupations (doctors, nurses, school teachers).

12.3.2.2.4 Female Overrepresentation in the Public Sector Tends to Reduce the Overall Gender Wage Gap

The question of wages in caring occupations is related to a well-known phenomenon of industrial gender segregation: women are overrepresented in the public sector. According to the OECD (2012), women accounted for about 58% of the total public sector workforce, and the public sector accounted for about 20% of overall employment in OECD countries. The overrepresentation of women in the public sector is explained by several nonexclusive factors (Blank, 1985): it offers traditionally female-dominated occupations (teaching, nursing, administrative tasks); it is often a family-friendly workplace with childcare, leave, and flexible hours, which are valuable to women; and employment security may attract more risk-averse workers (see Section 12.3.3.2).

Pay formation in the public sector is regulated (with variations across countries), and the public wage structure is more compressed than in the private sector (Gregory and Borland, 1999). As a result of this difference in wage structure, the gender wage gap is often smaller in public sector jobs than in the private sector (Arulampalam et al., 2007; Chatterji et al., 2011; Lucifora and Meurs, 2006). Another stylized fact is that the public–private sector pay gap is positive (with some exceptions, such as in Germany or Sweden; cf. Melly, 2005), particularly in the lower part of the wage distribution, but may be insignificant or negative at the top (Depalo et al., 2011). Because women are concentrated in the lower wage brackets, they are better off in the public sector (Meurs and Ponthieux, 2008); this is particularly true for the United Kingdom (Lucifora and Meurs, 2006) and Australia (Barón and Cobb-Clark, 2010). So, women's presence in the public sector tends to reduce the overall gender wage gap. However, this protective effect has

[48] It has been argued that this result comes from a positive selection of women in the labor market because only more highly qualified women stay on the labor market.

been receding with the trend of privatization since the 1980s and, more recently, with the budget crisis and its negative consequences on public wages (de Castro et al., 2013).

To sum up, horizontal gender segregation by occupation is often associated with a female wage penalty and is responsible for part of the gender wage gap in some countries (the United States), but this is not the case everywhere, nor is it the case in all female-dominated occupations. Labor market institutions and trade union density can reduce or reverse this penalty.

12.3.2.3 Vertical Segregation and the Glass Ceiling

As women gain education and experience, their relative absence in top jobs[49] is attracting growing attention (OECD, 2012) as a potential waste of human capital. More generally, vertical segregation—that is, the fact that women do not move up the career ladder at the same speed as their male counterparts—is considered today as the key factor for understanding the persistence of the gender wage gap.

12.3.2.3.1 Measuring the Gender Gap in Promotions and the Glass Ceiling

Women's disadvantaged position in the wage hierarchy can result from two different, nonexclusive mechanisms. The first is that women are promoted at lower rates than men. Since the publication by Albrecht et al. (2003), this has been called the "glass ceiling effect," evoking an invisible barrier between women and the highest positions in firms. A glass ceiling is suspected if the gap between the conditional wage distributions of men and women gets wider in the upper tail of the wage distribution. The second mechanism occurs when women are as likely as men to be promoted but gain less from their promotion than men do. This has been labeled the "sticky floor effect," following Booth et al. (2003)[50]; a sticky floor is suggested by a widening of the gender wage gap at the bottom of the wage distribution.

The detection of a glass ceiling effect, based on quantile regressions, has given rise to a large set of empirical work. The initial pattern of a continuous increase of the gender gap throughout the wage distribution and a sharp acceleration in the upper tail, as found in Sweden by Albrecht et al. (2003), was subsequently observed in many countries, but with differing intensities. The analysis of gender wage distributions in 11 European countries by Arulampalam et al. (2007) confirms the variety of patterns across countries and across sectors; this diversity is even more striking between the 26 European countries (Christofides et al., 2013). Kassenboehmer and Sinning (2014) examined changes in the US gender wage distributions over time, applying a decomposition based on the

[49] On average, across the OECD, only one-third of managers are female, with considerable variation across countries.

[50] Using a British household panel data set for 1991–1995, Booth et al. (2003) observed no differences between men and women in the likelihood of promotion, but rather smaller pay increases for women after promotion than for their male counterparts.

method described by Firpo et al. (2009). They confirmed the heterogeneity along the wage distribution, with greater narrowing of the gender wage gap at the bottom and small changes at the top. A glass ceiling effect is observed in the majority of countries, with the exception of Spain, but its magnitude varies substantially, depending on labor market institutions and the compression of national wage distributions. Countries with more "generous" work–family reconciliation policies have a lower unexplained wage gap at the bottom of the wage distribution and a wider one at the top—this is the case in Denmark and the Netherlands. Interestingly, a glass ceiling is observed in Spain only among educated workers (Amuedo-Dorantes and de la Rica, 2006; Del Río et al., 2011).

Gobillon et al. (2015) proposed another empirical tool for measuring the glass ceiling effect, based on a job assignment model. This measure is the probability ratio of getting a job for women and men at each rank of the wage ladder. An application to French full-time executives 40–45 years old in the private sector shows that the gender difference in the probability of getting a job increases along the wage ladder from 9% to 50%.

12.3.2.3.2 A Gender Difference in the Pace of Promotion Within Firms

We now turn to what happens inside the firm and the distribution of men and women across jobs. It has long been observed that pay rates tend to be equal in the same narrow occupational category within a firm (Blau, 1977, 2012; Groshen, 1991; Petersen and Morgan, 1995). This does not mean that there are no wage differences (and no discrimination) at the firm level, but rather that the observed gender wage differences at this level are mainly because of differences in the distribution of men and women in the wage hierarchy within the firm. This then raises the question of the extent to which gender differences in job-level assignment are due to labor supply factors, to firm discrimination, or to individual preferences.

To study precisely the internal mechanism of promotion requires access to personal data. These data are quite difficult to get, so there is a limited number of papers on gender differences in intrafirm mobility and job assignment. Overall, they conclude that there is a large part of job assignment that is not explained by observable individual productive characteristics. Malkiel and Malkiel (1973) studied the gender wage difference using a sample of professional employees in a single corporation for 4 years. When only individual variables are included in the Oaxaca–Blinder decomposition, a large part of the gender wage difference cannot be explained; adding the job level (defined in 13 categories) into the wage regressions makes the unexplained gender wage difference fall to a value close to zero. The authors showed that half the gender difference in job level cannot be explained, suggesting gender discrimination in job assignment.[51]

[51] Notice also that among the small and homogeneous group of US top executives, Bertrand and Hallock (2001) found that the huge gender pay gap (around 45%) is entirely explained by observable characteristics (age, seniority, firm size) and occupational segregation (women are less likely to be chief executive officers than men are). But this is not observed in every country. Smith et al. (2013) found a large unexplained gender gap in promotions to top positions in a cohort of Danish executives.

Pekkarinen and Vartiainen (2006) presented another approach to the glass ceiling effect by studying gender differences in the allocation of workers across jobs of different complexity among Finnish metalworkers. Women are less likely to be promoted than men, and women face a higher promotion threshold than men. Ransom and Oaxaca (2005) investigated intrafirm mobility and gender wage differences in a large grocery retailer in the United States over the 1976–1986 period. Importantly, for hourly wage workers, the wage rate in a given occupation is fixed for a given seniority by gender-neutral union contracts. Nevertheless, the employer had full control over job allocation and recruitment (in 1982, no woman held a top position). For hourly paid jobs (covered by union contracts), all gender difference in pay can be associated with the job assignment of employees, and women have less chance of promotion than men.[52] Dina et al. (2012) exploited two particularities of this case: there was no "pure discrimination"—union contracts were gender neutral—so the wage differential arose uniquely because of occupational segregation, and the firm was found guilty of discrimination, so the job assignment was proven to be discriminatory against women. The methodology developed in that paper extended the existing decompositions of the gender wage gap to isolate the contribution due to job segregation, even in the case where there is no overlap between men's and women's occupations.

12.3.2.3.3 Explaining the Gender Gap in Promotion

How can this gender gap in promotion within a firm be explained? Some studies argue that there is no unexplained difference once the exit rate has been taken into account; female executives exit the occupation at higher rates than men, and the executives who remain are promoted more quickly (Gayle et al., 2012). The promotion of young professionals also seems to be correlated with working longer hours (Gicheva, 2013), and women may be disadvantaged on account of their family constraints. These explanations of vertical segregation in terms of gender differences in labor market commitment are in line with human capital theory (Polachek, 1981). A model proposed by Lazear and Rosen (1990) formalizes the link between outside opportunities and wage careers. It assumes that wage differences are uniquely due to differences in promotions along a job ladder where top jobs require specific training. Men and women have the same distribution of labor market ability, but women have a higher ability in domestic work. Consequently, women have a higher probability of leaving the labor market, and, as a result, employers are reluctant to invest in their training and to promote them on an equal basis with men. Thus, a woman must have greater ability than a man to be promoted (and to be induced to stay at work by a higher wage). So their productive "advantage" in the labor

[52] When controlling for age, age squared, seniority, and seniority squared in a probit model, a male food clerk is more than six times more likely than a female food clerk to be promoted to a store-level salaried, managerial position.

market turns out to be a handicap in their wage career and prevents them from obtaining high-paid positions.[53]

This explanation does not, however, rule out the possible role of discrimination in job promotions (see the study on Austria by Winter-Ebner and Zweimuller, 1997, where a large part of gender differences in job positions remains unexplained even after taking into account expected job-separation probabilities). Statistical discrimination (Arrow, 1973; Phelps, 1972) is regularly advanced as a key notion for understanding inequality in careers and unequal access to higher positions. Models are based on workers' heterogeneity, imperfect information about their productive characteristics, and distinct distributions of unobserved productive characteristics between groups. This covers two situations: when women's level of ability is lower on average than men's (which is not true discrimination because workers are not equally productive) and when the ability levels are equal on average but the variance is larger within one group (see Aigner and Cain, 1977). The most common example of a case of statistical discrimination is when employers do not hire or promote women because they assume that women are more likely to interrupt their careers to have children and/or have a higher rate of absenteeism or be less available for late meetings and business travel because they are or will be mothers. This mechanism may be perpetuated by feedback effects: if women are systematically less often promoted than men because their employers anticipate career breaks, they may be less motivated to invest in their careers. They thus confirm their employers' preconceptions. The trouble is that such feedback effects make it difficult to distinguish between statistical discrimination—that is, unequal treatment due to employers' assumptions of labor supply behavior—and the human capital model of unequal promotions based on employees' work–family trade-off decisions.

These two lines of explanations both highlight the key role of career interruptions (effective or potential) in the gender gap in promotions. Other models complete these explanations based on family constraints by taking into account behavior at work and individual preferences on the gender composition of coworkers. This also implies that improvements in gender equality may be obtained by paying attention to the relationship between workers. Becker's model of taste discrimination was extended to unequal promotion by Baldwin et al. (2001), with the hypothesis that men have distaste for female supervision. Goldin (2013) presents a "pollution model" that combines the taste and statistical discrimination models: the entry of women into a male-dominated occupation is viewed as a negative signal on the value of this occupation and reduces the associated prestige. Consequently, male employees will oppose the recruitment of women to

[53] Note that one consequence of the higher productivity threshold to promotion is that women's average productivity in good jobs is higher than men's. In Lazear and Rosen's (1990) model, wages equal output in top jobs, so there is a female wage premium at this level, which is at odds with the empirical evidence, as underlined by Lazear and Rosen themselves.

protect their labor status. Empirical research also examines the role of interpersonal relationships within the firm to explain gender differences in promotion. Cannings and Montmarquette (1990) found that men use informal networks for career advice more often that women, whereas women rely more on formal bidding for promotion than men and wait longer for promotion.

12.3.3 Psychology, Social Norms, and the Gender Wage Gap

The persistence of the gender wage gap combined with the weakened explanatory power of standard human capital variables (education, experience) has led economists increasingly to integrate psychological factors and sociological approaches into their analysis of gender differences on the labor market (Bertrand, 2010). Psychosocial factors are used and tested to explain occupational segregation in the gendered choice of field of study, job, and sector. Also, because the gap in promotion to the top positions in the wage distribution is one of the most important gender inequality, it is not surprising that a large part of the psychological work applied to economics focuses on the reasons why women seem to be disadvantaged in labor market competition. The issue still being debated is the explanatory scope of these findings in the real world. These traits certainly play a role in slowing women's promotions, but their effect on the gender wage gap should not be overestimated.

12.3.3.1 Gender Differences in Risk Aversion and Competitiveness

Women's risk aversion and lack of competitiveness have been extensively studied in experiments (Croson and Gneezy, 2009; Eckel and Grossman, 2008), and these psychological traits are said to explain gender differences in earnings (see, e.g., Dohmen and Falk, 2011; Gneezy and Rustichini, 2004; Gneezy et al., 2009; Gneezy et al., 2003; Niederle and Vesterlund, 2007, 2011). Laboratory and field experiments concur that women are more risk averse and less competitive than men.

The reasons for these gender psychological differences are still being investigated in terms of the respective roles of culture (Gneezy et al., 2009 found that women in matrilineal societies are more competitive than men), social learning (Booth and Nolen, 2012a,b used variations in single-sex coeducational schooling and found that all-girl groups favor risk-taking attitudes), or biological factors (on the role of hormones see Wozniak et al., 2010). If psychological features are acquired, the next question is the age at which these gender differences in behavior appear, so experimental studies examine the attitude toward competition among the young (Dreber et al., 2011; Gneezy and Rustichini, 2004) or the very young (Sutter et al., 2013 extended the analysis to 3-year-old children). A gender gap in competitiveness is generally observed at young ages, except by Dreber et al. (2011), whose research is based on Swedish children, among whom no gender differences were observed. Again, the role of social environment, even in childhood, seems to be central in explaining gender differences in psychological traits.

A close and related question is the stability over time of these psychological traits. Risk aversion and the taste for competition often are considered as immutable when people reach adulthood. But new evidence suggests that it may still change over time and in different circumstances. To test this is not easy because it requires repeated information on individual preferences, but some recent empirical papers found that risk aversion is not time invariant: middle-aged people seem to be more risk averse than adolescents and elderly people (Tymula et al., 2013). Risk tolerance is also positively influenced by health (Hammitt et al., 2009), and unexpected negative shocks such as a disease change the attitude toward risk (Tison et al., 2012).

12.3.3.2 Occupational Segregation, Risk Aversion, and Gender Identity

Psychological approaches have been applied to the analysis of gender occupational segregation in many different ways. For instance, psychosocial traits such as self-esteem, impulsivity, or self-assessed intelligence may shape the choice of both study subjects and occupations (Antecol and Cobb-Clark, 2013). Risk aversion is also an important feature in the choice of occupation because it may lead to a trade-off in favor of more stable but less well-paid occupations (Bonin et al., 2007). Because the public sector offers more job security, it may attract more risk-averse workers (Clark and Postel-Vinay, 2009) and thus explain the overrepresentation of women in the public sector. The few empirical studies of this subject[54] confirm the influence of risk aversion on self-selection into the public sector (Bellante and Link, 1981 for the United States; Pfeifer, 2011 for Germany). The preference for team work may also explain industrial segregation. In laboratory experiments, Kuhn and Villeval (2013) found that women are significantly more likely than men to select team-based pay. They conclude that this preference may explain their overrepresentation in the nonprofit sector and helping occupations, where work often requires cooperative production (but provides few financial rewards).

However, the most widely used framework for understanding gender differences in choice of occupation is the model of gender identity developed by Akerlof and Kranton (2000, 2010). Their starting point is the idea that each individual is assigned a social category ("man" or "woman") associated with prescribed behaviors. Failing to comply with these prescriptions leads to disutility in oneself and in others (negative externalities). An action's payoff is then dependent on gender identity. So, men in female-dominated occupations (such as nursing) or women in male-dominated occupations (law) do not follow the behavior expected of their gender. They make their coworkers uncomfortable; in return, the latter may react negatively and not cooperate with them. These two negative payoffs (the disutility of deviating from social norms and the disutility brought by coworkers) may help to explain why women (or men) are reluctant about certain

[54] Few data sets provide a measure of individual risk aversion and the choice between public and private sectors.

occupations and why occupational segregation persists over time. This framework is extended from the choice of occupations on the labor market to the global question of the gender division of labor, particularly when the time of work within the household cannot be explained solely by specialization and also seems to be shaped by social norms (see Section 12.6).

12.3.3.3 The Impact of Women's Lack of Competitiveness on Wages

Two main psychological factors (reluctance to ask, lack of competitiveness) are put forward to explain women's disadvantage in promotion opportunities. Gender differences in the ability to negotiate have been studied by Babcock and Laschever (2003). In a US cohort of graduates from the same business school, they observed that men obtained a first job salary 7.6% higher than women; however, only 7% of women had tried to negotiate with their recruiter, compared with 57% of men. A large field experiment further clarified the difference between men and women in initiating wage bargaining (Leibbrandt and List, 2012). The authors collected 2500 responses from jobseekers to fictitious ads for administrative jobs in various cities in the United States; when the ads did not explicitly mention that wages were negotiable, men were more likely to negotiate than women, but if it was explicitly stated that wages were negotiable, this gender difference disappeared. Another important result of this testing is that male candidates preferred the rules of wage determination to be ambiguous.

In laboratory experiments, Gneezy et al. (2003) found that women are less effective than men in a competitive environment, particularly when they are in direct competition with men, but they perform equally well in noncompetitive environments. The findings of Niederle and Vesterlund (2007) corroborate this idea that women tend to avoid competition, unlike men (on aversion to competition and the choice of tournament see also Datta Gupta et al., 2013). The authors attribute these results to male overconfidence and gender differences in the taste for competition. This behavior reduces women's chances in competing for jobs or promotion, especially when they are in direct competition with men.

The question of whether experimental findings can be extrapolated to real life is discussed by Levitt and List (2007), and the contributions and limits of experimental designs to the understanding of gender differences in labor market outcomes are discussed by Azmat and Petrongolo (2014). Some studies affirm that the influence of women's aversion to competition in explaining gender differences in wage promotion should not be overestimated. Garrat et al. (2013) developed an interesting approach using the natural experiment of a footrace where the participants, both women and men, had to choose between two levels of competition (elite race and cash prize or "regular"). As expected, women were, on average, less interested than men in the elite race. But the very fastest women respond to financial incentives and were quite likely to try the elite race. It was in the midrange that the difference between women and men was the widest: in this range, men tended to overestimate their skills. The authors concluded that the economic consequences of aversion to competition for capable women are probably limited in the labor market. One example of discrepancy between behavior in laboratory experiments and behavior in the real world is given by Lavy (2012): in a case of performance-based pay for teachers, contrary to lab and field experiments, he did not find gender differences in performance when the financial rewards depended on a rank order tournament.

The last two examples are based on case studies, so it is difficult to extrapolate their significance to the whole economy. Manning and Saidi (2010) adopted a more general approach by using a large individual data set for Great Britain with detailed information on performance pay contracts. They tested the importance of differences in attitudes toward competition on the actual gender wage gap. They found that women are less likely than men to work on performance contracts, as predicted by experimental studies, but the final effect of performance pay on earnings is limited and does not differ much by gender. So, they concluded that the competition hypotheses do not provide a significant explanation of the observed gender pay gap. A similar conclusion (that a small part of the gender wage gap is explained by psychological factors) was obtained in a previous work on the gender gap in early career wage growth (Manning and Swaffield, 2008).

12.3.4 Family Constraints, Career Interruptions, and the Family Pay Gap

The more women participate in the labor market, the more they face the need to reconcile the constraints of motherhood and professional life. The unequal sharing of unpaid work and family responsibilities (see Section 12.6) probably contributes much more to gender income inequalities than women's lack of taste for competition. Fortin (2005) argued that culture, social norms, and inner conflict between work and family values may be considered a major cause of the slowdown (or, recently, the stagnation) of the gender convergence in pay.

Consequently, family situation and the composition of the household (particularly the number and age of children) are crucial to the gender wage gap. As Polachek (2004) argued, "the detrimental division of labor is at the root of almost all the wage gap" (p. 27). In the labor market, family characteristics have dissimilar consequences for women and men, even between women and men who are comparable in their productive characteristics. To what extent can women's adaptations to their family constraints—by choosing a family-friendly occupation with flexible hours, choosing a part-time job, or interrupting her employment—explain the average gender wage gap and differences in career? How large is the motherhood wage penalty between comparable women?

Men's wages also are affected by their family status, but in the opposite direction to women. Their earnings are not modified by the number of children they have (which is consistent with the fact that fatherhood does not change their work habits), but men enjoy a marriage premium, while there is no such positive wage difference for married women. There are two main explanations for this marriage premium: either married men are positively selected, or they are more productive thanks to specialization at home and the domestic chores done by their spouses. The earliest studies of this issue tried to distinguish between these two hypotheses; more recent studies focus on the expected changes in the marriage premium when the norm becomes a two-earner family, the benefits of specialization within marriage are receding, and the traditional type of married couple becomes less prevalent in OECD countries.

12.3.4.1 On the Demand Side: Lower Wages as a Result of the Inelasticity of Female Labor Supply

Because of their family obligations, (married) women often face multiple constraints in their choice of labor supply; this can give employers in monopsonic labor markets some market power. In this case, the employer exploits the inelasticity of the supply curve and pays workers below the competitive wage. The more inelastic the labor supply, the lower the wage (see Manning, 2003, on dynamic monopsony drawn from the search-theoretic framework of Burdett and Mortensen, 1998). In this approach, married women or mothers can suffer a wage penalty because they are less (or not at all) geographically mobile or because they need to work closer to their home to meet their domestic responsibilities. The female labor supply to a given firm may then be less elastic than that of men, and the employer may pay them lower wages than comparable men. Ransom and Oaxaca (2010) tested this framework in the case of a retail grocery store chain and found that the elasticity of labor supply to the firm differed between men and women; this difference was consistent with wage discrimination against women. Other studies from Germany (Hirsch et al., 2010) and Norway (Barth and Dale-Olsen, 2009) also found that women's labor supply is less elastic than men's and is linked with wage discrimination.

12.3.4.2 The Supply Side: Part-Time Work and Career Interruptions

Motherhood may also modify women's labor supply as a result of the gender division of labor within the household (see Section 12.6.3). Part-time paid work is a common way for mothers to reconcile work and childcare, and it may be considered a substitute for childcare facilities.[55] Part-time work usually leads to a pay penalty (in hourly earnings) compared to full-time work. In the United Kingdom, the penalty was about 25% and has widened over recent decades (Manning and Petrongolo, 2008). This penalty is partly due to individual characteristics (part-timers are less educated) and partly due to occupational segregation: women switching to part-time work often downgrade professionally (i.e., move to a lower-skilled job[56]). The risk of occupational downgrading is strongly

[55] In countries such as Australia, Germany, Japan, the Netherlands, and the United Kingdom, more than 40% of women work on a part-time basis (OECD, 2007).

[56] This phenomenon has been labeled as a "hidden brain drain" in a report by the UK Equal Opportunities Commission (2005).

dependent on the (lack of) part-time opportunities within a woman's current occupation (Connolly and Gregory, 2008).

Does this correspond to a willing choice by women, happy with this reduced time, making it possible to reconcile work and family? Or, does it rather correspond to a constraint on their labor supply? The answer depends greatly on the institutional context and on the social norms in the country considered. In three countries characterized by a high share of part-time work (Australia, the Netherlands, and the United Kingdom), Booth and Van Ours (2008, 2009, 2013) found that women report high levels of satisfaction with their reduced working hours, which are associated with a clear gender bias in the division of labor within the household. The case of the United Kingdom seems slightly different; women's satisfaction seems to be unaffected by their hours of work and remains "a puzzle."

Career interruptions are a radical way to meet family constraints but have strong and lasting negative effects on wages. Since the work of Mincer and Polachek (1974) and Mincer and Ofek (1982), employment interruptions are commonly responsible for gender differences in earnings because of the loss of experience and human capital depreciation (skills atrophy). Are the (negative) returns associated with career breaks different for men and women? Empirical evaluation is difficult for want of an appropriate data set with precise information on the duration of employed periods, time out of the labor market, and the reasons for career interruptions. It is therefore not surprising that the magnitude of the penalty associated with career breaks varies according to the study and methodology, but it is regularly found to be significantly negative. Kim and Polachek (1994) identified three possible biases in the estimates of penalties related to career interruptions: heterogeneity (unobserved characteristics such as motivation, influencing both labor market intermittency and earnings); endogeneity (low wages may explain intermittency and not the reverse); and selectivity. Taking into account these biases and using panel data (the Panel Study of Income Dynamics), they found a stronger effect of labor market intermittency on wages, but the effect was equivalent for men and women; the unexplained male–female wage gap was, therefore, very small. Albrecht et al. (1999) also analyzed the coefficients associated with the total time out using a rich Swedish data set that distinguishes different reasons for time out of the labor market (formal parental leave—an important component of the Swedish family policy—versus household time and other career interruptions). The penalty associated with parental leave for mothers was far smaller than that for fathers. The interpretation is that parental leave taken by the father is a rare event (compared with parental leave taken by mothers), so the signal was interpreted negatively by employers; for mothers, the negative coefficient reflects uniquely human capital depreciation. Finally, Hotchkiss and Pitts (2007) evaluated the total effect of labor market intermittency on gender wage differentials using a US sample of retirees with detailed information on their work history. They found that differences in intermittent employment represented 61% of the explained part, making the largest contribution to gender differences on observed characteristics[57]; but the sample was composed of retirees belonging to generations when labor market interruptions were very common.

12.3.4.3 The Family Pay Gap

Since the mid-1990s, increasing attention has been paid to the "family pay gap" or "motherhood wage gap," that is, the differential in hourly wage between women with and without children. This is because the wage dispersion between women with different family characteristics increased in some countries (the United States) while the gender wage gap was shrinking. As a result, marital status and children were making a larger contribution to the average gender wage gap (Korenman and Neumark, 1992; Waldfogel, 1998a), whereas the wages of childless women were closer to those of men. The family pay gap is generally negative, but its extent varies depending on the country, the period, and the methodology used to assess it.

Several nonexclusive explanations for the family pay gap have been advanced (see the detailed survey by Budig and England, 2001). First, working mothers are more likely than childless women to have spent time out of the labor market and, consequently, to have accumulated less human capital. In addition to the negative impact of career breaks on wages, having children to care for may influence mothers' allocation of effort (Becker, 1985). Mothers might limit their occupational choices to jobs and positions that are compatible with their family responsibilities—seeking "more convenient and less energy-intensive jobs" or just putting less effort into their work or being absent more often (Simonsen and Skipper, 2012, found some evidence of the role of absenteeism in the wage penalty in Denmark). These adjustments, under the form of working or having worked part-time jobs or in jobs or firms that are more "family friendly" (or simply closer to the home or school) may result in reduced work opportunities or, according to the theory of compensating differentials, lower pay to make up for better working conditions (Filer, 1985). Finally, mothers may be discriminated against if their employers assume that they are less productive and therefore avoid promoting them or putting them on fast tracks.

Empirical measurement of the effect of children on women's wages is not straightforward. One possible method is to include the number of children in a standard wage equation based on a cross-sectional data set; the motherhood penalty then corresponds to the estimated coefficient. However, this approach suffers from several biases, the most severe being the unobserved heterogeneity between mothers and nonmothers: less work-oriented women may have more children and cause spurious correlations of motherhood and wages. So, the most widely used method to measure the family pay gap is using panel regression models to control for cohort effects, individual unobserved heterogeneity, and, if possible, sample selection bias (working women are not a random

[57] However, the explained part represented only 30% of the total gender wage gap for this sample.

sample[58]). However, one problem with cohorts is that, by construction, one observes the results of past careers. Because there have been profound changes in education, job opportunities, and family policies over the past decades, the observed family pay gap may not be an accurate prediction for the next generations of women.[59]

Most of the evidence of the existence and extent of a family pay gap is based on Anglo-American research (Budig and England, 2001; Joshi et al., 1999; Lundberg and Rose, 2000; Viitanen, 2014; Waldfogel, 1998b). There are also some results for other countries, for example, Spain (Molina and Montuenga, 2009, using the European panel ECHP[60]) and Australia (Livermore et al., 2011, based on the Australian panel Hilda), and a few comparative studies: Harkness and Waldfogel (2003) studied seven industrialized countries; Davies and Pierre (2005) studied the European Union; Gangl and Ziefle (2009) studied the United States, the United Kingdom, and Germany; Sigle-Rushton and Waldfogel (2007) studied eight Western industrialized countries. Significant pay penalties associated with motherhood (around 2–10% for one child, 5–15% for two or more children) have regularly been found in all countries except the Nordic countries.

In the United Kingdom and the United States, the wage penalty is entirely explained by career interruptions, reduced work hours, and the concentration of mothers in low-paying part-time jobs (especially in the United Kingdom, where the family gap is particularly high). Joshi et al. (1999) found a penalty of 33% for the United Kingdom (comparing mothers and nonmothers at the age of 30), and Waldfogel (1998b) found a penalty of 20% (for women aged between 30 and 33 in 1991) in the United Kingdom and the United States. Interestingly, using a different method (propensity scores) on the same British cohort as Waldfogel (1998b), Viitanen (2014) found similar results for the United Kingdom and also found a long-lasting (but small) pay penalty 30 years after motherhood for the generation born in 1958 (the last wave observed was of women aged 50–51 in 2008–2009). For the United States, Kahn et al. (2014) found an attenuated effect of motherhood at older ages, except for women having three or more children, which is consistent with the work effort explanation (young children are more demanding than older ones). Anderson et al. (2003) proposed an alternative explanation; observing that medium-skilled mothers suffered a higher net wage penalty than low- and high-skilled mothers in the United States, they argued that medium-skilled workers face more time constraints (regular presence required during office hours, no possibility of taking work home), with negative effects on their wages.

[58] Variables used to correct for selectivity are generally household wealth and nonwage income (see, e.g., Datta Gupta and Smith, 2002).

[59] Another practical problem is the lack of national longitudinal data sets, which explains why most studies are done in the United States and the United Kingdom.

[60] The ECHP (European Community Household Panel) was a panel survey running from 1994 to 2001 that covered a wide range of topics concerning living conditions. It was replaced by the EU-SILC in 2003/2004.

A more recent debate concerns the relationship between the timing of births and the family pay gap, with the hypothesis that women postpone having children to accumulate work experience and build their careers; the fertility delay seems to be correlated with an increase in earnings (Caucutt et al., 2002).[61] The empirical difficulty is to identify the causal effect from delayed maternity to earnings and so to find convincing variables that influence the timing of motherhood and nothing else. Miller (2011) exploited biological fertility shocks in the US case and confirmed a positive effect of delaying the first birth on wages (as well as a flatter wage profile after motherhood).

There are few studies of continental Europe because of the lack of suitable data sets. One exception is Germany, where the GSOEP (German Socio-Economic Panel) allows studies comparable to those from the United States and the United Kingdom. A pure and high wage penalty against working mothers (i.e., an unexplained difference after taking into account labor market behavior) has been observed in Germany for cohorts born between 1955 and 1969 (Gangl and Ziefle, 2009). This finding was confirmed by Felfe (2012), who tested the influence of job amenities on the wage penalty in Germany. The motherhood wage gap can be partially explained by compensating wage differentials, but women who have neither changed jobs nor reduced their working hours still face a family wage penalty of 12%.

Nordic countries constitute a special case because the family gap is very limited (Harkness and Waldfogel, 2003, compared seven western countries including Finland and Sweden; Albrecht et al., 1999, studied Sweden; Datta Gupta and Smith, 2002, studied Denmark). One possible explanation of this singularity is that the generosity of Nordic maternity leave and family policies affects all women, mothers or not, since childless women are expected to become mothers and take parental leave; this leaves room for statistical discrimination against all women, however many children they have (Datta Gupta and Smith, 2002). Another important feature for understanding the small motherhood penalty in Nordic countries is the role played by self-selection into the public sector, which offers more favorable conditions to mothers.[62]

12.3.4.4 Opting Out: Do Highly Skilled Mothers Interrupt Their Careers?

The studies presented above generally found that the motherhood penalty increases with the education level (Anderson et al., 2002[63]), and that interruptions are less frequent

[61] However, the mechanism can be different for the access to the few top positions. Smith et al. (2013) found that having children at a young age increases the probability of being selected as chief executive officer among the pool of vice presidents. The interpretation could be that these women who were obliged to reconcile work and family have proven their high productive capacity.

[62] We will return to the adverse effect of Nordic family-friendly policies on women's wages.

[63] Using the US National Longitudinal Survey labor market experience of the Young Women 1968 cohort, they found that only highly skilled women suffer a wage penalty (much higher for those with two or more children), which is, for white mothers, entirely explained by years out of the workforce.

among highly skilled mothers, because this has a higher cost for them. However, the idea that highly educated women are increasingly "opting out" to care for their children as a consequence of "mother guilt" is receiving much more media coverage and has attracted the attention of economists as a possible effect of culture and norms on labor supply and a possible explanation for the small number of women in top jobs. To test this hypothesis, various studies have examined the careers of highly educated cohorts of women and their propensity to leave their job and the labor market, temporarily or permanently. No strong evidence of the opt-out revolution was found. On the contrary, Goldin and Katz (2008) even found evidence of a stronger labor market attachment in the most recent group of women from highly selective colleges (Harvard/Radcliffe classes) who graduated in 1970, 1980, and 1990, and no major change in fertility rates across cohorts (around 38% in each cohort were childless 15 years after graduating). Half of the mothers from these cohorts had not been out of the labor market for more than 6 months, and the most recent graduates had taken even less time off after childbirth.

However, highly skilled workers are heterogeneous with respect their work commitment. Focusing more specifically on women graduated from a top US business school between 1990 and 2006, Bertrand et al. (2010) found that MBA graduates seem to have more difficulty combining career and family than physicians, PhDs, or lawyers. This is consistent with a study by Herr and Wolfram (2009), who found that female graduates from Harvard are less likely to stay in the labor market if their workplace lacks a family-friendly policy. The phenomenon of opting out also seems to be more common among a specific group of women in highly visible job positions or working in particularly male-dominated occupations. The propensity of the media to focus on this group may explain the discrepancy between common perception and empirical evidence (Antecol, 2011). The second implication of these findings is that family constraints matter for the labor supply, even for some highly educated women. Pursuing these analyses, Goldin (2014) explains the heterogeneity among highly qualified women by differences in the allocation of working hours and the nonlinearity of earnings with respect to hours worked. People working in corporate, financial, and legal occupations are considered not substitutable and are required to work long hours (and are paid accordingly). Lack of flexibility makes the life of women with young children too difficult and incites them to leave their jobs. On the contrary, pharmacists are perceived as being interchangeable and can modulate their working time, so women are not disadvantaged in such occupations.[64]

12.3.4.5 On the Men's Side: A Marriage Premium

Family situation also influences men's wages, but in the opposite direction. There is generally no direct effect of children on men's wages, and there are few studies of this issue

[64] Notice that Goldin (2014) deduces that gender convergence in wage may come from a firm's changes in pay structure, with more linearity in paid hours of work to favor temporal flexibility.

(Loughran and Zissimopoulos, 2009), but a differential of 10–20% has been regularly observed in married men's earnings compared with unmarried men (mostly in the US literature), and there is an abundant literature to elucidate the reasons for this wage advantage.

The two main hypotheses tested are productivity and selection. More precisely, men's marriage premium may result from greater productivity resulting from more effort (Becker, 1974b) or more time spent at work (Korenman and Neumark, 1991), thanks to their wives being specialized in home production. Employers might also favor married men over unmarried men because they think they are more stable (Hill, 1979a). The marriage premium may also result from a selection mechanism: more productive men in the labor market are also more valued in the marriage market (Cornwell and Rupert, 1997; Nakosteen and Zimmer, 1997). Bonilla and Kiraly (2013) recently proposed a model in which the marriage premium results from search equilibrium in frictional labor and marriage markets, but they have not tested it empirically.

Empirical analyses provide mixed evidence on selection versus productivity hypotheses, which is not surprising given the endogeneity problem, the sensitivity of results to the methodology, the variety of identification strategies, and the fact that the two hypotheses are compatible. For the United States, Nakosteen and Zimmer (1997), observing earnings before the marriage, and Dougherty (2006), using panel data and taking into account the duration of marriage,[65] conclude there is a selection effect. On the contrary, Chun and Lee (2001) found no effect of selection and a positive effect of the degree of specialization within the household.[66] Mehay and Bowman (2005) came to a similar conclusion, exploiting personal data and observing performance reviews and promotions of US naval officers. Korenman and Neumark (1991) used a fixed-effects model and found that the marriage premium seems to be due to married workers being in better-paid positions within the company and receiving higher performance ratings from their supervisors, lending support to the productivity effect explanation. Ginther and Zavodny (2001) exploited shotgun weddings (marriage because of premarital conception) as a potential exogenous cause of marriage to identify the selection effect and concluded that selection does not explain the marriage premium, as did Antonovics and Town (2004), who used data on monozygotic twins. In addition to these results from the United States, Petersen et al. (2011) found an opposite result in Norway, where the marriage premium resulted from sorting into higher-paid occupations and occupation-establishment units before marriage. More important, the marriage premium seems to be decreasing over recent decades (Blackburn and Korenman, 1994), and

[65] Dougherty also found a small and temporary but positive wage premium for women, which is incompatible with the specialization hypothesis.

[66] Chun and Lee used a switching regression model with endogenous marital selection; the identifying variable was local marriage market tightness (the relative differences between male and female adults by state).

marriage may even negatively affect the wage growth of men in the United States for recent cohorts (born in 1979) when unobserved heterogeneity has been taken into account (Loughran and Zissimopoulos, 2009). This would be consistent with the idea that marriage—and dual-earner families—impose constraints on men's careers, too.

So far we have not considered the impact of changes in family status and the increasing share of cohabitation on the marriage premium. Does this premium differ according to the legal status of the couple? Datta Gupta et al. (2007) examined the situation in the United States and Denmark and found in both countries that marriage is a more selective state than cohabitation[67] and that it is therefore important to control for cohabitation when estimating the marriage premium. They also observed a small but negative effect of fatherhood on wages in Norway—the wage growth is lower for fathers than nonfathers—and explained it by the hypothesis that fathers may devote more time to childcare and less to training. For Germany, Barg and Beblo (2009) used a nonparametric matching model to differentiate between wage premiums for married and cohabiting men compared with single men. A wage premium exists for cohabiting and married men and seems to be due to positive selection into both marriage and cohabitation. Finally, Killewald (2013) considered different types of family (coresidential or not, biological father or not) in estimating the marriage premium and found a vanishing marriage premium in less normative family structures (unmarried fathers, nonresidential fathers, or stepfathers). Moreover, married residential fathers also received no statistically significant wage premium when their wives worked full time.

12.3.5 Institutions and Policies Matter

Almost all OECD countries have legislation ensuring equal pay for equal work regardless of gender (OECD, 2007). However, there is evidence of wage discrimination on the labor market, as proved by the persistence of an unexplained part of gender wage differentials. Policies against wage discrimination, and more generally against gender inequality at work, can be classified into two broad categories: those aiming to regulate firms' behavior (firm-targeted policies) and those aiming to change women's labor supply and facilitate the conciliation between work and family.

Policies seeking to change firms' hiring and promotion behavior do not seem to be very effective on the whole because they do not address the major source of inequality on the labor market, namely family constraints. Consequently, public policy, especially family-friendly policies, is now considered to be the major lever in reconciling work and family (OECD, 2007, 2012) and in reducing gender inequalities on the labor market. Public policies encouraging women's labor market participation also are considered important for economic growth and the fight against poverty and to maintain fertility rates in OECD countries. These public policies cover monetary incentives to work

[67] But there is also a wage premium for cohabitation.

(tax/benefit systems, childcare benefits) and work–family conciliation policies (parental leave, public childcare services). They are effective enough in increasing female labor market participation (with differences across countries according to the bundle of policies), but family-friendly policies that are too generous may have a "boomerang effect" and jeopardize women's careers. Ultimately, involving fathers in childcare through paternal leave incentives is increasingly viewed as a way to directly address the source of gender inequality, but empirical evidence is still rare and does not show any dramatic changes in childcare sharing.

12.3.5.1 Firm-Targeted Policies

In addition to enacting antidiscrimination legislation, changing the behavior of firms toward working women can be done in two opposite ways: first through the pressure of competition, and second by regulating the composition of the workforce (affirmative action, quotas). Both have been scrutinized in numerous works. In both cases, there is empirical evidence of their effects on gender inequality, which are generally positive but of a limited extent and sometimes dubious.

12.3.5.1.1 Does More Competition on the Product Market Reduce Wage Discrimination?

According to Becker's (1971) theory of taste discrimination, discrimination is costly for firms, and gender discrimination should spontaneously regress with increased competition: discriminatory employers cannot survive in a competitive market because they have to pay higher wages than their competitors to obtain their preferred type of worker. The policy recommendation is then simply to let the market clear discriminatory employers. In recent years, a number of empirical studies have tested this hypothesis by studying the evolution of the gender wage gap in contexts of increasing competition. This could be the result of deregulation in a given industry, as for the US banking industry (Black and Strahan, 2001); they found that men's wages decreased more than women's and that women's occupational status improved. Using a difference-in-difference approach, Heyman et al. (2013) also found that takeovers and product market competition have a positive (but limited) impact on the relative position of Swedish female employees. A more common way to study the effects of a competitive shock on the gender wage gap is to observe the consequences of the opening up of an economy to international trade. Examining the relationship between globalization and changes in the gender wage gap across industries in the United States, Black and Brainerd (2004) found that the decrease in the residual gender wage gap was faster in concentrated industries after the shock of competition than in already competitive industries. Similar findings were obtained for other developed countries. According to Klein et al. (2010), exporting firms exhibit less wage discrimination than nonexporting firms in Germany. Meng and Meurs (2004) also found a smaller unexplained gender wage gap among firms more exposed to competition in France and Australia. These papers are based on national case studies.

Zweimüller et al. (2008) gave a more general view by adding to the meta-analysis described above (Weichselbaumer and Winter-Ebner, 2005) a proxy for the degree of competition in each country (the "Economic Freedom Index"). They observed a negative correlation between competitive markets and gender wage gap residuals. They acknowledged, however, that they could not tell whether these reduced gender wage gaps resulted from wage increases for women or wage decreases for men, and the effect on employment was not considered. The general conclusion of these diverse studies is that competition actually reduces the gender wage gap (generally by reducing men's wages), but the effect is not strong enough to profoundly change firms' discriminatory behavior.

12.3.5.1.2 Affirmative Action and Quotas: Their Effect on Gender Equality

Another means of reducing discrimination on the labor market that has been investigated is directly regulating firms' labor demand and compelling them to observe certain rules in their workforce composition. The famous US affirmative action programs of the early 1907s[68] fall into this category of antibias policy and have been extensively discussed in the literature (Fryer and Loury, 2005), mainly based on Coate and Loury's (1993) theoretical framework, in which it is argued that positive discrimination policies do not systematically eliminate negative stereotypes. A large strand of research on the effect of this policy is devoted to the employment of racial minorities in the United States,[69] but it also has been extended to women's employment. Leonard (1989) concluded that affirmative action has been more effective in increasing job opportunities for minorities than for white women, whose employment would probably have increased without the program. However, Holzer and Neumark (2000) have a more positive assessment of affirmative action policy; they argue that the firms adopting this program change their process of hiring and training, reducing statistical discrimination, and attract more minority and female job applicants. Eberts and Stone (1985) focused on the effect of Equal Employment Opportunity Commission on promotion in the education sector in the 1970s and also concluded that discrimination against female teachers had decreased. Finally, experimental evidence concludes that affirmative action may increase women's willingness to compete (Niederle et al., 2013) and their self-confidence (Villeval, 2012).

[68] Title VII of the Civil Rights Act of 1964 bars private employment discrimination on the basis of race, color, religion, sex, or national origin. The policy was enforced with the Equal Employment Act of 1972; since then, all federal contractors have been required to maintain written affirmative action plans for women.

[69] One strand of empirical literature studies access to self-employment status by US black minorities (Fairlie, 1999) and changes in access through this affirmative action policy on subcontractors. Blanchflower and Wainwright (2005) concluded there are few changes despite the existence if various programs, and Chatterji et al. (2013) concluded there is a more positive effect of small business set-aside programs on minorities' employment than affirmative action programs.

More recently, attention has been centered on gender quota policies in top management and their effect on gender inequality. This is consistent with the analysis that the persistent imbalance in top positions is now the most important mark of the inequality between men and women on the labor market. It has led to public policies imposing an increase in the proportion of women on corporate boards in many OECD countries (Pande and Ford, 2012), with the notable exception of the United States, which is very hostile to this kind of regulation. Norway is a leader in this field, requiring 40% women on corporate boards since 2003, but now it is also the case in France, Italy, the Netherlands, and Belgium, which have mandatory quotas from 30% to 40%. A European directive imposing a quota of 30% on corporate boards for all member states in 2015, and 40% by 2020, was adopted in 2012.

The percentage of women on boards of directors is usually analyzed from the point of view of firms' financial performance, with mixed evidence, more often positive (for the United States see Carter et al., 2007; Erhardt et al., 2003; Miller and del Carmen Triana, 2009; Shrader et al., 1997, via a positive effect on innovation; see also Smith et al., 2005, on a panel of Danish firms; Adams et al., 2011 found a positive reaction of shareholders to female director appointees in Australian firms). The problem with this empirical evidence is that variations in the percentage of women on boards of directors are endogenous, so causal effects are difficult to prove. In Norway, firms were obliged by law to change their board composition, so it offers an interesting natural experiment. Exploiting this case, Ahern and Dittmar (2012) found negative effects of the shock on board composition: stock prices and firms' performances (measured by Tobin's *q*) both decreased. They interpreted these results as a consequence of the lack of experience of new female directors replacing old and experienced male directors. This suggests that the negative effect may be transitory if women accede to positions enabling them to acquire managerial skills.

This raises the question of the effect of gender quotas in boards of directors on human resource practices. Does the reinforced presence of women at the top trickle down to the lower grades? The expected mechanism behind this hypothesis is double: women at the top may directly favor the promotion of women and/or act as positive role models. There are still few econometric studies testing the effects of top managers on the female proportion of executives because of the dearth of suitable data. The few studies of US panel data found a positive association between an increase in the proportion of women on boards of directors or among top managers and a reduced gender wage gap (Cohen and Huffman, 2007), better pay for executive women (Bell, 2005), or a higher share of women among top executives (Matsa and Miller, 2011; and for Norway, Matsa and Miller, 2013; Kurtulus and Tomaskovic-Devey, 2012). However, a positive association between the proportion of women among top executives and positive outcomes for women is not observed everywhere; for example, in Denmark, Smith et al. (2013) did not find that having more women in the management board favors the promotion of women. Therefore, the link between the proportion of women involved in the hiring process and a break in the glass ceiling effect is probably not as straightforward as these findings may suggest. Bagues and Esteve-Volart (2010)[70] showed that a higher proportion of women on a recruitment committee plays in favor of male candidates, whose relative quality is overestimated. Other studies (based on correspondence testing) did not find gender difference in the recruiters' bias in favor of men for hiring in science research positions (Moss-Racusin et al., 2012). This sheds doubt on the efficacy of quota policies for helping women in their careers. To sum up these mixed findings, the presence of women in top positions does not guarantee that promotions will be fair or that the lack of women at the top will quickly dissipate.

The mitigated results of voluntary policies to equalize career prospects between men and women are partly due to the timing of childbearing. Some policy programs directly address the difficulty women have in combining a career and childbearing ("the overlap between biological and tenure clocks"), especially in academia. The "Stop the Clock" policy, which delays promotion review under certain circumstances, was introduced in 1971 at Stanford University and is now widely adopted in US academic institutions. Manchester et al. (2010) found no empirical evidence of the effectiveness of this policy and even proposes that Stop the Clock policies may exacerbate the gender pay gap in academia. This last example suggests that the effects of equal opportunity programs are limited when they do not directly address issues related to labor supply.

12.3.5.2 Public Policy and Women's Participation in the Labor Market

The particularities of the labor supply of (married) women have been extensively studied in labor economics (Heckman, 1974; Killingsworth and Heckman, 1986). Basically, it depends on the choice between market activities (and the corresponding wages) and household production. The mechanisms behind this choice, however, seem to have evolved dramatically. As underlined in the first part of this section, the increase in the female labor supply since the 1970s has been outstanding, and this trend was mainly due to changes in the behavior of married women with young children; the traditional pattern of withdrawal from the labor market to take care of children has receded in recent decades. Olivetti (2006) explained this impressive change by the relative increase in (married) women's returns to experience. As a result, work interruptions are penalized more highly, especially for young women in their early careers. These changes shifted the elasticity of married women's labor supply in two directions (at least for 1980–2000 in the US case): their wage elasticity decreased by half and became quite similar to men's, and their labor supply became less responsive to their husbands' wages (Blau and Kahn, 2007).

The general increase in total hours worked by women may be the result of greater participation in the labor market (extensive margin) and/or of an increase in the intensity

[70] These findings are based on evidence provided by Spanish public examination for judiciary occupations (magistrates). The 150,000 candidates were randomly assigned to juries of varying gender composition.

of work on the job (intensive margin). In two companion papers, Blundell et al. (2011a,b) decomposed the evolution of total hours worked in three countries (the United States, France, and the United Kingdom) into extensive and intensive margins for the period 1975–2008. One striking result is that while the extensive labor supply for married women with children is similar in the three countries, the intensive margin differs: US married women increased their mean annual hours of work, contrary to women in France and the United Kingdom, who tended to reduce their hours worked. These differences between industrialized countries are related to differences in institutions and policies, for instance, whether part-time work for mothers is encouraged.

We focus here on the two main policy tools influencing the female labor supply: earnings taxation and family-friendly policies (cf. OECD, 2007). First, we briefly sum up the debates about the response of the labor supply to income tax. To what extent does the current fiscal structure determine the labor supply of men and women? Second, maternal and parental leave, childcare subsidies, and family-friendly public policies in general have been expanding over the past few decades in OECD countries, but at different speeds in different countries. A growing corpus of studies details their impact on the labor behavior of mothers. The positive impact of these policies on female labor participation is widely recognized, but various studies highlight the possible negative effects of some of these policies—especially parental leave—on female wage careers. The parental leave taken by fathers has more recently been viewed as a possible lever to change attitudes at work and to improve equal gender opportunities, but the (still) scarce empirical evidence on this issue does not really support this view.

12.3.5.2.1 Marriage, Taxation, and the Labor Force Supply

Does it pay for the second earner in a household to work in the labor market? The answer greatly depends on the tax system for families and on the average effective tax rate and the marginal effective tax rate (METR).[71] The OECD (2012, Chapter 4) provides an overview of the METR for test cases in OECD countries. The designed test cases are the transition from inactivity to part-time employment, then from part-time to full-time employment, for couples with two young children (aged 4 and 6). The financial incentive to take a part-time job in single-earner couples is very low in 16 of 29 countries (particularly Sweden, Switzerland, Ireland, and Norway, where the METR is >100%), whereas the incentive to transition to full-time work is strong. In both cases (transition to part-time or to full-time work), the incentives are stronger for dual-earner couples.

These test cases assess the effect of national tax systems on family income without differentiating between incentives within the household. In fact, income tax may be applied

[71] The average effective tax rate is the proportion of gross earnings lost by taking up employment (extensive margin). The METR measures the proportion of any increase in earnings lost to taxation and/or benefit income reductions for those already working (intensive margin).

to individuals or to households, and the choice between joint taxation and individual taxation is the subject of considerable theoretical debate, with strong implications for female labor force participation. A large majority of authors support the individual tax unit through the Ramsey (1927) rule: an optimal tax system ensures that individuals with high labor supply elasticity face a lower marginal tax rate than those with low elasticity. Applied to dual-earner married couples and when the income tax is progressive, this means that it is preferable to tax individual earnings rather than the household income (Boskin, 1975); in the case of individual taxation, the secondary workers of the household, whose labor supply elasticity is high, are taxed at a lower marginal rate than the primary workers, whose labor supply elasticity is low. Following this reasoning, the US system, in which husbands and wives have been allowed to pool their income and file a joint return[72] since 1948, is generally considered not optimal (Rosen, 1977) because it discourages second earners (in practice, wives) from working.

The Boskin and Sheshinski (1983) model of taxation of married couples relies on individual and cross–labor supply elasticities and the joint distribution of wage rates to design optimal taxation; their numerical example, based on US parameters, suggests that the optimal rate for husbands would be approximately twice that for wives. One of the difficulties with this scheme is that individual taxation counteracts the objective of redistribution: if the government values redistribution, two married women with the same labor income ought not to be treated identically if their husbands' incomes are different. Kleven et al. (2009) studied how the tax rate of one individual should vary with the earnings of the spouse, taking into account the objective of redistribution. They propose a sophisticated model of optimal income taxation of couples where the second earner has to decide whether to work (a binary choice). The optimal tax formula is a function of labor supply elasticity, the redistributive tastes of the government, and the distribution of earning abilities and work costs among the population. As a result, the optimal tax (or subsidy) on secondary earnings decreases with primary earnings and converges to zero asymptotically.

The idea that a tax system based on individuals rather than households is more efficient has been challenged by Piggott and Whalley (1996) by adding household production to the model of optimal taxation. In this case, the optimal tax design should not distort the input of family members in household production. The underlying intuition is that moving from an individual to a household-based income tax reduces the supply of secondary work but increases time for household production and leisure and can improve welfare.[73] Apps and Rees (1999, 2007) disagree with this model. They also include household production in their optimal taxation model and find, contrary to Piggott

[72] In this case, the taxation is calculated as if they had each earned one-half of the income.

[73] Note that in this model, the household is treated as a single optimizing agent.

and Whalley (1996), that it is still optimal to tax individuals separately, as in the Boskin and Sheshinski (1983) model.

Another lively debate of the optimal taxation of married couples concerns the idea of gender-based taxation (GBT), whereby women have a lower tax rate because of their higher elasticity of labor supply. Alesina et al. (2011) argued that such a reform would increase the bargaining power of women within the family, endogenize labor supply elasticities, and narrow the gender labor elasticity gap. Their idea is to replace the various policies in favor of women, such as quotas, affirmative action, or subsidized childcare facilities, with this tax advantage for women. Saint-Paul (2008) strongly disagreed on the grounds that it is discriminatory and GBT would be perceived as unfair because it does away with equality before the law. He argued in favor of a gender-neutral system (the second earner not necessarily being a woman) with the same rate of taxes on individual earnings. Guner et al. (2012b) brought an empirical perspective to this debate. They tested the introduction of GBT in the US context and found that it would improve welfare, but welfare gains would be higher if the US tax system were replaced by a proportional, gender-neutral income tax.

There are relatively few empirical studies of the effect of fiscal policies on the actual female labor supply. Guner et al. (2012a) used a dynamic model with heterogeneity and quantified the effects of two possible tax reforms in the United States: a proportional income tax and a reform in which married individuals file taxes separately. Their model indicates that the effect on married women's work hours is stronger in the second case—labor force participation increases more than twice as much as it does under a proportional income tax reform (+10% and +5%, respectively). The effect is even more pronounced for married women with children, with men's hours being nearly constant in both cases. Kabátek et al. (2014) studied the French case, which is characterized by the joint taxation of married couples (but not for cohabiting couples at the time of the survey), using an individualized data set with information on each member of the couple. Once again, a simulation of separate taxation concluded that the labor supply of wives increases, but only to a limited extent.

Japan offers an interesting case because the tax system is such that the primary earner (usually the husband) has a spousal deduction (Allowance of Spouse) if his spouse earns income below a threshold level. As a result, an increase in the wife's earnings decreases the amount deductible from the husband's income. Akabayashi (2006) exploited this particularity to assess the labor supply response of women and found that it is a strong disincentive to women's labor supply because they value more the loss of the spousal deduction than the potential gain of their own paid work. Consequently, the intrahousehold allocation of labor supply is inefficient.

As in Japan, Italian tax policy does not encourage women to participate in the labor market: a tax credit for dependents is paid to the main earner of the family, which encourages one-earner families. Figari (2011) studied the potential effects of abolishing the existing tax credit and replacing it by in-work benefits, inspired by the UK Working Tax Credit, which is an income supplement for all individuals working 16 or more hours a week. The first scenario envisaged is an in-work family-based benefit; the second is individual based. Using EUROMOD (the tax benefit microsimulation model for the European Union),[74] Figari found an increase in the labor supply of women in both cases; the larger effect was obtained with the individual-based benefit, and the labor supply changes are particularly marked for the poorest households.

Figari et al. (2007) proposed comprehensive comparative research on the effects of the tax and benefits system on the labor supply of married women for nine European countries. As in the case study for Italy, the simulations were based on the EUROMOD model, which allows for calculations on the basis of the actual social and demographic characteristics of each national population. The first aim of this comparative work was to quantify the difference in independent income brought into households by male and female partners in couples and to measure how far the gap is closed by the national tax and benefits system.[75] In all nine countries, the average share of couples' pretax benefit income received by women is far less than half, with the lowest in Greece (18%), mainly because of its low female rate of participation, and the highest in Finland (37%). But the case of women with a higher income than their partners is not so rare (around 25% of couples in Finland and the United Kingdom). The tax benefit system reduces within-couple income inequality on average in all countries: more so in Austria, Finland, the United Kingdom, and France (countries where non-means-tested benefits[76] have an equalizing effect), and less so in Greece (the system makes little difference to the women's share before and after tax benefit). The equalizing effect is small in Germany and Portugal. The second aim of this study was to assess the incentives of women in couples with an earning partner to increase their earnings by working more or at a higher wage rate (intensive margins) or by taking on paid work themselves (extensive margins). As expected, comparative analysis indicates that joint taxation (France, Germany, and Portugal) is the system with the most important negative impact on incentives for the second earner to work (assessed based on marginal effective tax rates).

12.3.5.2.2 Family Policy Instruments and Labor Market Participation

After this overview of the debates about the effect of the tax system on female labor market participation, we now turn to public family policy and its impact on the female labor

[74] Note that the paper is based on a unitary model of household behavior (pooled income constraint).

[75] Independent incomes are allocated to the member of the household who earns or receives them. Benefits received by an individual to compensate for individual risks (maternity, unemployment, disability) are allocated to the individual concerned. Family or child benefits are divided equally between the two partners. The convention is questionable, but the authors assume that it is more transparent to adopt this assumption in the absence of information about the actual sharing.

[76] These benefits are composed of parental benefits, unemployment benefits, and child and family benefits (equally shared by assumption).

supply. Family policy instruments cover a wide variety of programs (OECD, 2013; Thévenon, 2013). Typically, they combine three kinds of measures: entitlements to maternity and parental leave after the birth of a child, which provide employment protection and often are covered by public income support; childcare services adapted to the working hours of parents with young children; and a tax benefit system that contains incentives to work.

All countries have developed these policies over recent decades, but to different extents and in different combinations, so it is not easy to obtain an overall picture of the bundle of policies. One convenient way to classify types of family public policy is to use the Esping-Andersen (1999) framework. Broadly speaking, countries can be categorized into three groups (Thévenon, 2011, 2013). Nordic countries (Denmark, Finland, Iceland, Norway, and Sweden) favor the model of the dual-earner family. These countries are characterized by comprehensive support for working parents with very young children. They offer generous birth-related leave conditions for working parents and provide childcare and services out of school hours. English-speaking countries (Ireland and the United Kingdom, the United States, Australia, Canada, and New Zealand) provide less generous in-kind support to working parents with very young children and more cash benefits, often targeted to low-income families. Continental countries form a heterogeneous group: southern European countries provide limited support for working families, but France is a particular case, providing strong support for working women to combine work and family. Interestingly, the type of welfare state regime (which can also be categorized as sociodemocratic, liberal, and conservative in the order of presentation above) also shapes the division of labor at home (Geist, 2005; see Section 12.6 on unpaid work in the household).

12.3.5.2.2.1 A Positive Effect of "Family-Friendly" Policies on Female Labor Market Participation The expansion of "family-friendly" policies in Europe (and in the other OECD countries, except for the United States) over the past two decades has largely contributed to the increase in female labor force participation, contrary to the recent stagnation of the US participation rate[77] (Blau and Kahn, 2013). The diversity of policies within Europe has attracted numerous studies seeking to evaluate their effect on female labor participation in a comparative framework. In general, macroeconometric analyses of OECD panel countries found a positive effect of family-friendly policies (especially when combined with female education, low unemployment, and favorable cultural norms) on female labor market participation; the exception is child benefits, which tend to reduce female participation because of an income effect (Jaumotte, 2003). Thévenon (2013) developed a macroeconometric analysis of the responses of female labor participation to changes in policies for 18 OECD countries from 1980 to 2007. His results confirm the key role of childcare services as a driver of female

[77] The US female participation rate has nevertheless stabilized at quite a high level (75%).

participation in the countries' labor forces. Taking into account complementarities with other policies, he showed that this effect is enhanced in countries with high levels of employment protection, longer paid leave, and other measures supporting working mothers. Using the European panel ECHP, Del Boca et al. (2009) found that part-time work opportunities (when well paid), childcare, optional parental leave, and child allowances are a greater determinant of the participation choices of women with lower levels of education. They also concluded that differences in social policies across European countries are responsible for a large part of the differences in female labor market participation across these countries. Gutierrez-Domenech (2005) studied the influence of national policies on a precise event in work life—women's transition from employment to nonemployment after a first birth—and compared five European countries over 1973–1993. During this period, Spanish mothers increased their probability of employment after giving birth, contrary to West Germany. The author points out the shift towards a separate taxation system, the increase in education and part-time employment to explain this divergence across countries.

12.3.5.2.2.2 Maternal and Parental Leaves: What Effect Do These Have on the Participation Rate? Maternity and parental leaves (generally taken by the mother) are one key component of family policy in all OECD countries (Thévenon and Solaz, 2013). However, national policies differ in their objectives and in the resources allocated. A first group of countries (Nordic countries) favor short, well-paid leave as a way to preserve mothers' work attachment; other countries consider parental leave as a substitute for formal childcare services and propose long leaves at a flat rate (if any); the US case is particular because it was one of the few countries that had no national maternity leave policy until 1993 (since then, 12 weeks' unpaid leave have been granted under the Family and Medical Leave Act).

How do women react to maternity leave entitlement with regard to their work commitment? The general findings are that maternity leave encourages labor market participation before pregnancy to qualify for leave benefits and reinforces work attachment and return to the former employer (Ruhm, 1998; Waldfogel, 1998a). But the empirical problem in studying the effect of maternity leave on mothers' labor supply is to find exogenous variations in taking maternity leave. The identification strategy usually exploits either an institutional difference across countries on a parameter of the policy considered or a sudden change because of a reform in a given country. The US case has been largely studied because, despite the absence of national coverage, it was possible to get differentiated coverage through state policy, unionized firms, or voluntary employer provisions. These studies then compare the behavior of women with and without maternity leave entitlement and measure their respective probabilities of returning to paid work.[78]

[78] There is still a selection problem, however, because this advantage is not offered randomly by firms.

Women entitled to maternity leave are more likely to take leave up to the maximum period allowed but return more quickly after 12 weeks than mothers with no entitlement. They are also more likely to return to their previous employer after childbirth (Berger and Waldfogel, 2004)[79] and to receive a wage premium that offsets the negative effects of having children (Waldfogel, 1998b). Positive effects in terms of job continuity were observed by Baker and Milligan (2008), who exploited the variability of maternal leave eligibility across Canadian provinces, and by Joshi et al. (1996) in work with the British 1958 birth cohort, which was the first generation to be offered statutory maternity leave; the gains of maternity leave and other family-friendly policies were unevenly distributed and more pronounced for educated women.

Parental leave policy (after maternity leave) has complex effects on the probability of returning to work after childbirth. Two statutory components—the length of job protection and payment during the leave period—produce opposite effects on the return to work. A longer period of job protection increases the probability of returning to work; paid leave increases the probability of taking the leave and then staying at home (Pronzato, 2009). So, the overall effect of parental leave varies according to the generosity of the conditions. Lalive and Zweimüller (2009) provided causal evidence of the impact of parental leave on employment and earnings[80] using two opposite changes in Austrian parental leave policy. The 1990 reform increased the duration of parental leave from 1 to 2 years, whereas the 1996 reform reduced it to 18 months (there was no change in the benefit paid during the parental leave, equal to 40% of net median female earnings). They found that extended leave led to significant short-run reductions in employment and earnings and delayed the return to work, even after the parental leave had expired. But they did not observe any negative effects of the parental leave extension on earnings 5 years after the birth, even for highly qualified mothers (contrary to a general view that longer parental leave has a negative effect on women's careers; see Section 12.3.5.2.3). Dustmann and Schonberg (2012) used a change in maternity leave coverage in Germany[81]; as in the work by Lalive and Zweimüller, they found that the extension of leave actually delays mothers' return to work.

12.3.5.2.2.3 The Effects of the Costs of Childcare on Mother Labor Supply We have already mentioned the general consensus that access to childcare is crucial to enabling parents to be in paid employment. The costs of childcare may be a serious barrier against labor market participation. When they are taken into account, the financial

[79] Notice that Klerman and Leibowitz (1999) argue that the effect of the Family and Medical Leave Act (1993) on job continuity is small and that the new law did not change the practice.

[80] Fertility responses to Austrian parental leave reforms are also thoroughly examined by Lalive and Zweimüller (2009). We do not report their results on this issue, which are beyond the scope of this section.

[81] Their study focused on children's long-term outcomes after a longer period of staying with their mothers.

incentive to work decreases, especially for lower-income families.[82] The disincentive effect is particularly strong in Anglophone countries, Japan, Israel, and Switzerland (OECD, 2012). Consequently, lower childcare costs increase female employment, but estimating the price elasticity of childcare presents serious methodological difficulties (Blau, 2003). Natural experiments offer a convenient way to identify the effects of childcare subsidies on labor supply. Baker et al. (2005) studied a new childcare policy of reduced fees in Quebec, Canada, and found a large and positive impact on the labor supply of mothers with preschool children.[83] On the contrary, such an effect was not found in Sweden after a considerable reduction in childcare prices (Lundin et al., 2008). The interpretation here is that highly subsidized childcare already existed before the reform, so further reductions had limited impact (not significant). Some studies underline that the effect on labor supply also depends on availability (see Kreyenfeld and Hank, 2000, for Germany) and on the quality of childcare (see Hansen et al., 2006, for the United Kingdom). Finally, out-of-school care is also an important factor in reconciling full-time work with motherhood when children are in primary school, but this public policy is not yet very developed in OECD countries.

12.3.5.2.3 The Boomerang Effect of Family-Friendly Policies

The evidence that family-friendly policies boost the female labor participation rate seems to be solidly established by numerous empirical studies. However, this does not imply any improvement in women's position in the labor market or any reduction in gender wage inequality. Blau and Kahn (2013) completed their observation of the relative advantage of OECD countries over the United States in female labor market participation (see above) with a relative disadvantage in the quality of work: US women are more likely than women in other OECD countries to work full time and to be managers or professionals.

The most controversial family policy is long parental leave, which is regularly associated with a wage penalty for mothers in various countries (Ruhm, 1998, describes nine countries in a seminal work; Misra et al., 2011, describe European countries; Phipps et al., 2001b, describe Canada; Beblo et al., 2009, describe Germany; Lequien, 2012, describes France). Although parental leave preserves the ties between women and their employers and thereby encourages women's participation in the labor market, it reduces their commitment to a career.

These negative effects of family policy seem predominantly to affect highly qualified workers through a "boomerang effect" (Datta Gupta et al., 2006, 2008). The extension of family-friendly schemes[84] seems to have a negative effect on the wages of women

[82] The advantage of reduced childcare costs for low-income families does not make up for limited income gains.

[83] But the authors also found negative effects of this policy on children's health and behavior.

[84] In the Danish case, this factor is reinforced by female segregation in the public sector, where family-friendly policies are the most generous. This partly explains the observed stagnation (or even a tendency to increase) of the gender wage gap (Datta Gupta et al., 2006).

(with children) because it leads to a higher probability of long career interruptions and penalizes their careers, particularly for the most highly qualified. The same conclusion was reached by Mandel and Semyonov (2005, 2006) in their comparative analysis of industrialized countries: the welfare state facilitates women's access to the labor force but not to managerial occupations. More generally, family-friendly policies may deter women's (mothers or not) career progression (see Albrecht et al., 2003; Arulampalam et al., 2007; Christofides et al., 2013). They also attract disproportionate numbers of women to the public sector, which offers more favorable working conditions at the expense of their wage career (Simonsen and Skipper, 2006). Datta Gupta et al. (2008) concluded that "extensive family-friendly schemes may even have created a 'system-based glass ceiling' hindering women's career progression" in Nordic countries.

12.3.5.2.4 Does Paternity Leave Change Parents' Behavior?

Paternity leave has recently attracted attention as a policy that directly addresses the main cause of gender inequality on the labor market, namely the time devoted to childcare by mothers. The expected effects are twofold: first, to shorten maternal leave by sharing parental leave; second, to change mentalities in the hope of obtaining a fairer division of housework over the long term (Huerta et al., 2013; OECD, 2012, Chapter 18). The finding by Nepomnyaschy and Waldfogel (2007) that the length of paternity leave has a positive influence on fathers' involvement in childcare constitutes an argument in this direction. However, the country studied was the United States, where there is no paternity leave policy, even though a large proportion of fathers take some (unpaid) time off work after a birth, so it could be argued that the fathers who take longer paternity leave are a selected group.

Contrary to the US case, European countries have adopted various measures in favor of paternal leave in the past two decades. The Nordic countries (Norway, Sweden, Finland, and Iceland) are the most advanced in this field, with well-paid paternity leave and a strong incentive to take it because part of the total parental leave is lost if not taken by the father. For instance, Iceland is often exemplified for its system of 3 + 3 + 3, meaning that a third of the parental leave is reserved for the father (13 weeks), a third is reserved for the mother, and the last third can be taken by either the mother or the father (in practice, usually by the mother).

What are the actual changes in fathers' behavior induced by this policy? Highly educated fathers are more likely to take parental leave (Sundström and Duvander, 2002); large parental leave benefits and longer duration of exclusive parental leave for fathers seem to be positively correlated with fathers' childcare time (see Haas and Hwang, 2008, for Sweden; Boll et al., 2013, for a cross-country comparison based on time-use data). But these findings are descriptive and do not establish a causal effect of paternal leave on fathers' behavior. Ekberg et al. (2013) took advantage of a natural experiment in Sweden (the 1995 "daddy-month" reform offers 1 month of parental leave for the fathers of children born after January 1) to study the short- and long-term effects on fathers' behavior. They observed a strong short-term effect of the incentives on male parental leave but no behavioral effects over the long term (measured by sick days taken to look after sick children younger than 8 years old). It is possible that this persistency in behavior was due to the limited scope of the observed change (only 1 month). A similar pattern was observed in Germany; the natural experience exploited by Kluve and Tamm (2013) was the introduction of more generous parental leave benefit and 2 "daddy months" in 2007 (the "Elterngeld reform"). Like Ekberg et al., they found no significant changes in the time fathers devoted to childcare among those subject to the reform (measured by the fathers' relative contribution to childcare during the children's first year). Generous paternal leave attracts fathers and so has an effect on the duration of maternity leave, but it has little effect on childcare sharing.

To sum up the findings on the gender wage gap, the most salient fact is that the gender wage has stopped decreasing in OECD countries since the late 1990s, after a sharp decline in the 1970s and 1980s. The reasons for this decline have been clearly identified: rising education levels and increasing participation in the labor market have brought women's human capital closer to men's. The reasons for the recent stagnation are nowhere near as clear cut. From a descriptive point of view, men's and women's occupational structures differ considerably, and women are less successful than men in their careers. A huge number of studies have tested a large range of sophisticated explanations that are not exclusive and are difficult to rank in terms of explanatory power. First, firms may discriminate against women, and testing, natural experiments, and econometric studies show that women are not treated equally in the hiring process or in promotions. Second, women's psychological features may impede them from being as aggressive as their competitors or from choosing certain occupations. Once again, there are laboratory experiments proving the existence of gender differences in risk aversion and competitiveness, but the external validity of psychological explanations for the gender wage gap is still open to question. Finally, family constraints and the issue of children and housework appear to be central to understanding the gender wage gap: young mothers are penalized in their careers, especially when they are highly qualified; women may choose family-friendly places at the expense of their careers; if firms think that childless women may become mothers in the future, they may be reluctant to train and promote them. As a result, national public family policies play a major role in explaining national differences in the gender gap. The new paradox is that a generous system in favor of parents of young children has positive effects on female participation but a negative impact on their careers.

12.4. THE CASE OF SELF-EMPLOYMENT

Self-employment is a two-faced concept in labor economics. On the one hand, it is considered the basis of economic dynamism and the soul of capitalism. In line with Schumpeter's theory, the entrepreneur is the prominent figure of capitalism: an individual capable of perceiving economic opportunities, combining various resources, and taking risks. The belief that small business is essential to the growth of the capitalist economy has policy implications, especially in Europe. One of the priorities defined by the Lisbon Strategy (March 2000) was to foster start-ups. More recently (January 2013), the European Commission proposed the Entrepreneurship 2020 Action Plan "to reignite the entrepreneurial spirit in Europe" and to address the structural, administrative, and cultural obstacles to entrepreneurship.

On the other hand, self-employment is presented today as a way of mitigating the human consequences of the economic crisis by diminishing the unemployment rate (even artificially). More generally, self-employment is considered a route out of poverty for unemployed or marginal populations. The flexibility allowed by self-employment is also seen as a convenient way to accommodate constraints on the labor market: it allows flexible hours and schedules, can often be done at home, and so offers an intermediate situation between inactivity and salaried work. It is therefore not surprising that the European plan in favor of entrepreneurship suggests that policy should aim "to promote specific actions for reaching out to groups that are [. . .] under-represented within the entrepreneurial population" (OECD 2013, p. 4): migrants, seniors, the unemployed, young people—and women.

Gender studies on self-employment lie at the crossroads between these two approaches. The gender gap in self-employment is analyzed in terms of the characteristics required to become an entrepreneur. Since Cantillon (1755) and Knight (1921), the entrepreneur has been viewed as a risk-taker or bearer of uncertainty; to this risk attitude, Lazear (2005) added balanced skills and varied backgrounds as personal characteristics that influence the choice of entrepreneurship. Thus many studies test gender gaps in human capital, risk aversion, and self-confidence as the main explanatory factors for the lack of women in small business. Once women are self-employed, however, the analysis focuses on this situation as a means for women to reconcile work and family. Studies of women's performance compared with that of men explain the observed differences, mainly in working hours and, more generally, in the business objectives: profit for men, work–life balance for women. Like the gender wage gap, gender differences in earnings then tend to be related to differences in labor supply and family constraints.

In this section, we present the statistics available on self-employment rates in OECD countries,[85] showing that women are less likely to be self-employed than men in all OECD countries. Why do self-employment rates for women differ from those for men? A growing strand of research examines the role played by psychological

[85] We do not consider female self-employment in developing countries. There is a lot of debate about microcredits and self-business as an efficient way for women to gain economic independence, but the issues are too far removed from the scope of this chapter, which is centered on gender inequalities in OECD countries.

characteristics (risk aversion, self-confidence, and perception of economic opportunities) in explaining this gap. Then we turn to studies of differences in the way men and women run businesses. Some studies examine the possibility that women face discrimination in the access to credit. But most analyses are centered on gender differences in family constraints as the main reason for differing performance between men and women.

12.4.1 Stylized Facts

12.4.1.1 The Fuzzy Scope of Self-Employment

Self-employment covers a wide and heterogeneous range of economic activities: traditional farmers, regulated professions in law or health, small business, start-ups, and so on. To measure it, the OECD has adopted a comprehensive definition: "Self-employment is a form of employment in which people work in their own business, farm or professional practice and receive some economic benefit for their work, such as wages, profits, in-kind benefits or family gain (for family workers). Volunteer work is excluded from this definition" (OECD, 2013, p. 32). Self-employed people can work on their own (i.e., own-account self-employment) or have employees. Business owners are excluded if they are not involved in the day-to-day operations of the business activity.

As Ahmad and Seymour (2008) emphasized, entrepreneurship should not be confused with the above measure of self-employment. Entrepreneurial activity is defined as "the enterprising human action in pursuit of the generation of value, through the creation or expansion of economic activity, by identifying and exploiting new products, processes or markets" (OECD, 2007). Consequently, entrepreneurship is not limited to small business units, but includes innovative large companies. In practice, studies often confuse self-employment and entrepreneurship. This has consequences on gender analyses, which tend to confuse the relative dearth of self-employed women with the lack of female entrepreneurship.[86]

In practice, comparable statistics on self-employment are based on national labor force surveys. The Global Entrepreneurship Monitor (GEM) measures entrepreneurship activities through annual household surveys of the adult population in 54 countries and includes a set of questions on motivations and aspirations. GEM is also the main source

[86] Another statistical problem is the inclusion in self-employment statistics of people who are falsely self-employed ("bogus self-employment" or "economically dependent workers"). This corresponds to an employment arrangement in which the worker has only one customer and is registered with tax authorities as a self-employed worker rather than an employee on the employer's payroll. The extent of bogus self-employment is, of course, difficult to assess. According to Oostveen et al. (2013), it is marginal in the 27 European Union countries, representing less than 1% of the workforce. It seems more important in certain countries such as Italy and the Czech Republic (Geissler 2012; Oostveen et al., 2013), where it may represent more than 10% of the self-employed (2% of the employed workforce). There are no significant differences between men and women in the incidence of this form of "self-employment."

of information on nascent entrepreneurship (businesses started less than 42 months and more than 3 months ago).

12.4.1.2 The Gender Gap in Self-Employment

In all developed countries the self-employment rate—the proportion of self-employed people relative to all employed people—is lower for women than for men, whatever the average level of self-employment.

In the 27 European Union countries in 2011, the self-employment rate was 9.7% for women and approximately the double for men (18.3%); the average rate was 15%, with great variations across countries (OECD, 2013). Some countries, such as Denmark, France, and Germany, have a much lower female self-employment rate (around 5%) than the EU average, whereas others, such as Italy and Greece, have much higher rates (around 20%). There are large contrasts in the evolution of these figures within each country. Over the 2000–2011 period, 13 EU countries experienced an increase in female self-employment rates; the most significant increases were observed in Slovakia, the Czech Republic, and the Netherlands. The other 13 EU countries saw a decline (or a stagnation in Latvia), and the most significant declines were in Lithuania, Portugal, and Romania (OECD, 2013).

The United States has a relatively low rate of self-employment, equal to 11% in 2009 (based on Bureau of Labor Statistics figures in Hipple, 2010) and stable since 2003.[87] The female self-employment rate is, as expected, nearly half the male rate (8% and 14%, respectively), and the gap is particularly large in incorporated self-employment (2% and 5%, respectively).

In all countries, gender segregation by industry in self-employment seems similar to the segregation in salaried employment. Women are relatively absent in manufacturing and building, which account for approximately 25% of male self-employment. Self-employed women are more likely to work in consumer-orientated services (health, social work, arts, education). These sectors represent around 40% of female self-employment (OECD, 2013).

12.4.2 Why So Few Women in Self-Employment?

The lower propensity of women to start businesses is well documented (see Allen and Langowitz, 2013; Koellinger et al., 2013; Langowitz and Minniti, 2007; OECD 2013, Chapter 2; for the most recent references based on the GEM database; Anderson and Wadensjo, 2008 on the Swedish case; Furdas and Kohn, 2010, on the German case). Once a business is started, in general there are no significant differences in survival rates

[87] This stability results from the combined impact of a decline in unincorporated self-employment and an increase in the rate of incorporated self-employment.

across genders when economic characteristics are taken into account (for the United States, see, e.g., Perry, 2002).

Why do women start fewer businesses than men? Gender differences in observed human capital are not sufficient to explain the gender gap in self-employment (Minniti and Nardone, 2007; Wagner, 2007). Furdas and Kohn (2010), using a detailed individual data set on start-up activity in Germany, decomposed the probability of starting a business by sociodemographic variables and personality traits. They found that men opt for start-ups more often than women with comparable sociodemographic variables (age, education, professional status, region, immigrant background, and family environment). Much of the gender difference is explained by the distribution among women of personality traits (risk tolerance, openness, emotional stability, creativity, need for achievement, etc.) that are less favorable to entrepreneurship.

A large strand of literature has sought to explain the lack of self-employed women in terms of psychological and social factors, which can be classified in three broad categories: risk aversion, perception of economic opportunities, and preferences for self-employment. According to the model developed by Kihlstrom and Laffont (1979), in which individuals make their occupational decision by comparing the risky returns of entrepreneurship with the nonrisky wage in the competitive labor market,[88] gender differences in risk aversion are a natural candidate for explaining the gender gap in self-employment. Risk aversion may also play a role by discouraging access to financial resources to start a business; using the English Household Survey of Entrepreneurship 2003, Sena et al. (2012) found that women are more reluctant than men to borrow from banks, and this gender difference in the search for external funding reduces their probability of becoming self-employed. In addition to risk aversion, perception of opportunities (see Kirzner, 1979), self-confidence, and knowing other entrepreneurs are crucial characteristics for starting a business. Consequently, the lack of self-confidence among women, the gender difference in perception of the economic environment (women tend to perceive themselves and the entrepreneurial environment in a less favorable light than men do), and the higher fear of failure lower the propensity to start a business (Langowitz and Minniti, 2007; Minniti and Nardone, 2007; Koellinger et al., 2013; Wagner, 2007). In addition, gender stereotypes (Gupta et al., 2009) and preferences for self-employment may contribute to the gender gap in self-employment transition and also have given rise to a large number of studies combining psychology and economics. For instance, Verheul et al. (2011), using a representative data set of more than 8000 individuals in 29 developed countries, distinguish two stages in the process of starting a business: the general preferences for self-employment—whatever the actual work status of the individuals

[88] For instance, Ekelund et al. (2005) tested the role of risk aversion on self-employment using a Finnish psychometric data set and found a sizeable negative effect of this psychological feature on the probability of being self-employed.

surveyed—and the actual involvement in self-employment. They find that a lower inclination for entrepreneurship among women explains a large part of the observed differences in the self-employment rate and conclude that general preferences appear to be the key factor behind the low self-employment rate of women.

However, some recent studies have argued that the importance of social and psychological factors and gender stereotypes should not be overestimated, and that the decision to be self-employed also depends on economic opportunities for both women and men. A macroeconometric analysis of longitudinal UK data, estimating the extent to which women are influenced by economic factors, shows that economic factors (gross domestic product [GDP], interest rates, house prices, employed and self-employed incomes) influence equally the self-employment choices of both women and men (Saridakis et al., 2013). The impact of economic factors on the decision to be self-employed also has been studied using microanalysis by comparing the behavior of individuals with similar human capital. Generally, the behavioral difference between men and women in the propensity to become self-employed may be strongly mitigated when the population is homogeneous. Analyses restricted to professional and managerial self-employment indicate that women choose self-employment to pursue a careerist model and to obtain work autonomy as men do (for the United States see Budig, 2006a). Leoni and Falk (2010), who analyzed the propensity to be self-employed among Austrian university graduates, find that age and the field of studies explain two-thirds of the gender gap observed in self-employment. When the group studied is limited to medical graduates, they no longer observe any gap between men and women in the propensity to be self-employed.[89] But it could be argued that medical graduates are a self-selected group that attracts women prepared to exercise as licensed professionals.

12.4.3 Self-Employed Women: Family Constraints and Gaps in Working Hours and Earnings

Given the lower propensity of women to start businesses, what factors might encourage women to become entrepreneurs? One motivation for women to become self-employed could be to avoid the gender wage discrimination they face on the labor market[90] and to be rewarded according to their productive characteristics. Evidence, however, is not very convincing: empirical work tends to find a small unexplained part in the gender earnings gap of self-employed workers, once productive characteristics and work hours have been controlled for (for the United States see Clain, 2000; for Canada see Leung, 2006).[91] But

[89] In general, however, women still have a lower probability of being self-employed in other fields of study.

[90] Note that self-employed women may suffer from consumer discrimination if the customers do not like being served by women.

[91] One exception in this set of studies is Lechmann and Schnabel (2012), who obtained a larger unexplained part for the self-employed than for salaried workers in the German case.

despite these results, which suggest that women have higher returns on their human capital when they are self-employed, the role of gender wage discrimination in the transition toward self-employment is not firmly proved (Leung, 2006; see Williams, 2012, for a set of European countries).

On the contrary, there is a general consensus that familial responsibilities play a dominant role in women's choice of self-employment and that this status is often chosen by (married) women as a way to achieve a better work–life balance, given their family constraints (Carr, 1996). The family status and the presence of young children is therefore a key variable explaining the choice of self-employment because the flexibility in working hours allows parents to reduce the cost of childcare (Connelly, 1992). Empirical studies have regularly found a positive influence of fertility on women's selection into self-employment (Boden, 1999; Macpherson, 1988). Women with young children are more likely than men to declare flexibility of schedule and child constraints as main reasons for becoming self-employed (for the United States, see Boden, 1999). Therefore self-employment seems to be a close substitute for part-time work and labor market inactivity, especially in countries where there are few public childcare services (see Georgellis and Wall, 2004 on the US case). To be covered by their husband's health insurance is also a positive factor of women's self-employment, especially in the US case (Devine, 1994). Lombard (2001) presents a careful econometric study of the choice of self-employment over wage/salary employment by married women in the United States. The women's requirements in terms of monetary (wages) and nonmonetary job attributes (flexibility and a nonstandard work week) were estimated in a first stage, and these estimates, together with husbands' health insurance, were used to explain the self-employment decision. The conclusion is that these three factors—the relative earnings potential when self-employed, the demand for flexibility and a nonstandard work week, and the husband's health insurance—positively affect the probability of women being self-employed.

Another way to demonstrate the key role played by work–family time in self-employment is to compare the use of time by self-employed people and employees. According to a study based on the Australian Time Use survey (Craig et al., 2012), self-employed mothers spend more time on domestic work and childcare and their paid work hours are shorter than those of salaried mothers, whereas the fathers' time does not vary much across employment types. Some studies also have found positive nonmonetary outcomes for women (mothers) who have opted for self-employment: self-employed married women report greater job satisfaction, less occupational burnout, and less negative spillover from work to home (based on a 1997 US survey; Hundley, 2001).

Given these differences in work commitment and working hours, it is therefore not surprising that self-employed women are regularly found to perform worse than men. The raw gender earnings gap (defined as the difference between male and female average self-employment incomes divided by the male average self-employment income) is around 35% for OECD countries as a whole, with large variations between countries.

Portugal and Poland (countries where farms are still important) have the largest gap (60%), Nordic countries (Iceland, Sweden, Denmark) the smallest (10%), and the US and most European countries are at around 40% (OECD, 2013, gender data portal).[92] But, as mentioned above, the unexplained part of this gap is small: in most countries this gender difference in earnings generally vanishes quite entirely when structural differences (human capital, financial capital, sectors of activity), and particularly women's reduced working hours and their commitment to children and housework, are taken into account (for the United States see Budig, 2006b; Fairlie and Robb, 2009; Marshall and Flaig, 2014; see Du Rietz and Henrekson, 2000, for Sweden; see Young and Wallace, 2009, for Canadian lawyers). Hundley (2000) analyzes this male/female earnings gap in self-employment in terms of their different reasons for being self-employed (compared with salaried work). The starting point is that self-employment offers a broader range of options of work organization than salaried work: self-employment is not bounded by minimum work hours (and corresponding minimum pay), nor by maximum hours, contrary to salaried work. Combined with the gender division of labor (see Section 12.6), this implies that self-employed married women will do more hours of housework than salaried women (and less market work), and conversely self-employed men will do less hours of housework than salaried men (and more market work). As a result, the gender earnings gap is greater in self-employment than in salaried work. Empirical analysis based on the US Panel Study of Income Dynamics survey confirms that marriage, family size, and hours of housework have a negative effect on self-employed women's earnings and a positive effect on those of self-employed men. Hundley (2000) also suggests that this pattern may operate as a barrier to entry into business by ambitious women, as they may fear that self-employment will oblige them to do the lion's share of the housework.

12.4.4 Are Women Discriminated Against in Their Access to Credit?

Another well-documented stylized fact on gender differences in entrepreneurship is that self-employed women make less use of bank loans than men do. Are women discriminated against by banks, experiencing more credit denials or higher interest rates than men? Or do they simply apply for less credit? This issue has been largely studied in the United States (Blanchard et al., 2008; Blanchflower et al., 2003); it is regularly found that women do not face discrimination in terms of access to loans, unlike ethnic or racial groups. Asiedu et al. (2012) confirmed these results using the latest Survey of Small Business Finances in 2003. The 2003 database allows them to extend the analysis to loan renewals (and not only new loans). Loan renewals, compared with new loans, are

[92] There are specific methodological difficulties in collecting data on self-employment earnings: how to distinguish between the return on capital and the payment for hours worked, how to treat negative earnings, the bias due to incomplete declarations, etc. (see Section 12.2). However, these biases are assumed to affect men's and women's earnings equally.

interesting because there is less information asymmetry, so unexplained differences in denial rates (with white men as the reference group) can be attributed to discrimination with more certainty. After controlling for selection bias and nonlinearity (i.e., the possible different impact of some of the control variables according to the level of applicants' credentials), Asiedu et al. found discrimination against ethnic minorities, but not against (white) women; they even found that women tend to pay a slightly lower interest rate than white males.

Then the limited use of banking credit may stem from gender differences in behavior. Although there may be no gender differences in the granting of bank loans, differences in the perception and expectation of their banking relationship may negatively affect women's demand for credit. Saparito et al. (2013) studied gender differences in the perception of banking relationships; their research was based on a survey of 696 matched pairs of business owner/managers and bank managers and that included qualitative variables. They found that gender influences the quality of the relation between business owner and banker: the male pairs (a male firm owner and a male bank manager) have the highest level of trust and satisfaction with credit access, and the female pairs have the lowest level of these indicators. Interestingly, an influence of gender is also observed in women's access to angel capital. In general, women are less likely to seek angel financing, although they have an equal probability of obtaining funding from this source; they are also more likely to seek and obtain financing from women angels (Becker-Blease and Sohl, 2007).

However, the role of behavioral variables should not be overestimated in explaining gender differences in access to credit; the differences in structural characteristics explain a large part of this gap. Using Survey of Small Business Finances over two decades, Cole and Mehran (2009) summarized the major characteristics of female-owned firms as follows. These firms are smaller, younger, more likely to be in retail trade and business services, and more likely to be organized as proprietorships rather than corporations. Female owners are younger, less educated, and less experienced than male owners. Econometric analysis indicates that these structural differences explain why female-owned firms are less likely to obtain credit than male-owned firms.

The above findings on the absence of gender banking discrimination are for the US case. There is no certainty that the result is valid in other OECD countries. In fact, Alesina et al. (2013) offer a counterexample in the case of Italy, where the share of self-employment is high and a lot of microfirms are owned by women. They find robust evidence that women pay more for credit than men, even after taking into account a large number of characteristics relating to the type of business, the borrower, and the structure of the credit market. The disadvantage is higher in male-dominated industries. The authors suggest two explanations for this result: a taste-based discrimination against women, as Italian society still has very traditional values concerning the place of women in society, and asking for a loan could be considered unsuitable, and a lower ability to negotiate good deals with banks because women are more averse to competition and more reluctant to ask.

To sum up, some features already highlighted in the analysis of the gender wage gap also seem to be relevant to gender differences in self-employment, namely the choice of field of study, industrial segregation, and (once again) family constraints and the need for flexible work hours to reconcile childcare and professional activity. The possibility of combining work with staying at home seems to be contradictory to the relative scarcity of self-employed women. Gender differences in risk aversion seem to provide a solid basis for explaining this paradox, but their importance should not be overestimated. Finally, self-employment is a very heterogeneous field, with huge differences in types of business (how can one compare the behavior of lawyers and farmers?) and across countries, and for which data are still too limited (compared with those for wages) to provide a comprehensive picture of gender inequalities.

12.5. THE GENDER GAP IN PENSIONS

Pensions are the main source of retirement income for the elderly, and it is well known that older women receive far less individual pension income than older men. All countries have adopted various mechanisms to take into account the wages lost by women because of their family constraints and to reduce their risk of poverty in old age. However, before 1990, the gender pension gap was a blind spot, virtually invisible in the economic literature (Ginn, 2001), and it has only recently attracted attention (see Jefferson, 2009, for a survey; Folbre et al., 2005). There were two reasons for this invisibility. First, because pensions were based on past earnings, their less advantageous careers automatically led to smaller pensions for women, and the gender pension gap was viewed as a simple outcome of this fact. Second, it was considered that nearly all women were married, and married women shared the benefits from their husband's entitlements, and this occurred even after his death through survivor's pension schemes. In other words, the hypothesis of income pooling within the family was explicitly assumed for retiree couples and there was no point in studying gender inequality. Consequently, the gender aspect of pensions was covered from the perspective of the living standard of widows and spinsters, as the people most at risk of poverty. The policy issue was that of a safety net for old people and poverty (mainly women poverty) and/or the more or less generous way to transfer pension rights to the survivor (wife).

The recent growing interest in the gender pension gap stems from three concurrent social evolutions. First, the end of the male breadwinner model in the 1960s has led to the current massive entry into the retirement system of generations of women with their own past work lives and their own pension entitlements. The usual analyses of the gender wage gap can then be transposed to the gender pension gap, with similar issues. Should we expect a narrowing of the gender pension gap similar to that observed for the gender wage gap? To what extent do part-time, incomplete careers, occupational segregation, wage discrimination, and so on have adverse consequences on the level of individual pensions? Second, changes in family history with the decline in marriages, the rise in divorces, and the increase in cohabitation and celibacy have a direct impact on the gender pension gap. The typical figure of a single retired woman is now less likely to be a widow and more often a divorcee or spinster. This change calls into question the rationale of survivor's pension schemes inherited from the past, when there was only one (male) breadwinner.

Last but not least, population aging and the corresponding increase in the dependency ratio are exerting pressure on retirement schemes in all countries. The general trend in reforms is to reduce the level of pensions and to align pension benefits more closely with career trajectories. A pure contributory system—defined as a system where the total sum of pensions received by a pensioner is proportional to the discounted sum of contributions made during his or her working life—creates gender inequalities because of the previous gender inequalities in careers. The more the pensions system is based on contributions, the more it is disadvantageous to women. However, all countries have kept or expanded diverse mechanisms to counterbalance these (negative) gendered effects, such as noncontributory pensions, the use of unisex mortality tables to calculate contributions, and benefits or additional family rights.

In this section, we present the (few) statistics on the gender pension gap. We will see that it is even more difficult to obtain comparable international figures of this indicator than it is for the gender wage gap. Then we discuss the main mechanisms that determine the actual level of pensions and the extent to which the same rules may have a different impact on men's and women's pensions. Finally, we turn to family history and its consequences on the gender pension gap. The main issue here is how to compensate for unpaid work such as caring for children or other family members and the emerging transition toward family rights and away from spouse's rights.

12.5.1 Sparse and Noncomparable Statistics

There is general agreement on the existence of a considerable gender pension gap, but there are no harmonized international statistics to compare the situation between countries. Some figures are available by country, but the way they are calculated is heavily dependent on the specificities of each national pension scheme.

One of the problems in obtaining comparable figures, which is also encountered for the gender wage gap, is the question of which population is considered. Is it limited to private pension recipients or not? Are social security pensions included or not? But there are also problems related to the variety of national pension schemes and the way data on pensions are collected. Is the gender pension gap measured at the household level, with information on the composition of earnings (direct pensions, derived pensions, savings income) or based on individual administrative data without information on the composition of the household? Finally, there is a difficulty specific to the gender issue. Unlike wages, which are defined on an individual basis and are a remuneration for hours worked, pensions may include rights derived from the spouse (pensions paid to the spouse or ex-spouse, survivors' pensions) or from the family situation (family rights).

There are very few comparable international statistics, even fewer than for the gender wage gap. One exception is the recent document published by the European Commission (2013) presenting comparable tables based on the EU-SILC 2010 database. To the best of our knowledge, these are the only published statistics comparing countries available in 2014 and, by construction, the United States is not included in the comparison. The population surveyed is defined as people over 65 years old not living in collective households[93] and receiving a pension. Because of the way the data are collected, the first "two pillars" of pensions—state pensions, based on societal solidarity and pay-as-you-go financing, and occupational pensions (OPs), based on occupational solidarity and prefunding—are not distinguished; only the third pillar (individual pension schemes) can be isolated. Another particularity of these figures is that survivor's pensions are mixed with other pensions.

According to this study, the average gender pension gap[94] in Europe was 39% in 2009, approximately twice the average gender wage gap for these countries.[95] Germany, the United Kingdom, and the Netherlands have a gender pension gap of more than 40%, whereas it is less than 25% in Nordic countries (Finland, Denmark), and Eastern European countries (Figure 12.6).

For the United States, scarce figures are given separately for the social security system (the first pillar in the United States) and for the private system. In both cases there is a gender pension gap, equal to 35% for the social security system (which includes spouse's benefits) and stagnant over the past 50 years, despite the increase in the labor force attachment of women. This stagnation is mainly due to the US spousal benefits system: additional social security earnings have no effect on social security benefits because they replace the spousal entitlement (see below). The picture is different for private pensions: between 1978 and 2000, the ratio of women's to men's benefits increased from 0.23 to 0.29 (Even and Macpherson, 2004).

An additional difficulty in obtaining pertinent figures is that the population of retirees is far from homogenous in terms of work history and accumulated entitlements, so the average pension is not very informative, particularly for women: recent cohorts have had

[93] This excludes people living in nursing homes, who represent a sizeable percentage of old people.

[94] The average gap is equal to 1 minus the average women's pension/average men's pension.

[95] The comparison should be treated with caution because the average gender wage gap is measured at a given point in time, whereas the average gender pension gap is based on the cumulated gender gap for the whole working life.

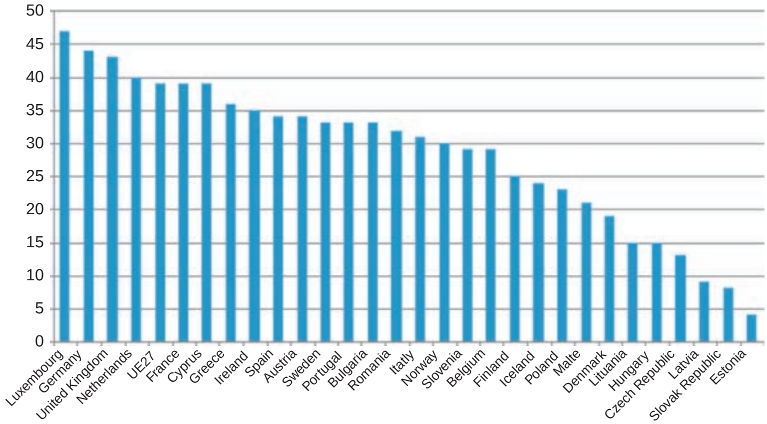


Figure 12.6 Average gender pension gap in Europe (pensioners aged over 65 years). *Source: Bettio et al., 2013 – Based on EU-SILC UDB data 2010.*

a longer working life than the older cohorts, for whom the past norm was the one-earner family. This implies a narrowing of the gender pension gap, reflecting the dramatic growth in female participation in the labor market. For instance, in France in 2008, women's own pensions were 44% of men's for the 1924–1928 cohorts and 56% for the 1939–1943 cohorts (Andrieux and Chantel, 2011). However, the picture is not really so clear cut and depends on the viewpoint one adopts. Because the SILC statistics on pensions are not limited to personal entitlement pensions but include survivors' pensions, the gender gap in total pensions by cohort does not indicate narrowing (Bettio et al., 2013). Comparing cohorts older than 80 years with the cohort aged 65–79, it can be seen that the younger groups face considerably wider gender pension gaps (41% vs. 33%). This surprising result is due to a combination of large proportions of widows who receive survivors' pensions among the eldest women and selection effects (life expectancy is higher for former qualified workers, so pension benefits are relatively higher and the group is more homogeneous). Consequently, changes in women's labor market participation result in attenuated changes in benefits for women; the benefits derived from spousal rights are replaced by benefits linked to individual careers.

12.5.2 The Gendered Effects of Pension Regulations

The individual gender pension gap mirrors work life inequality: women's lower labor force participation, more part-time work, more career interruptions, and lower pay all have consequences on their pension entitlements and levels. Most studies (Even and Macpherson, 2004, for the United States; Bonnet and Geraci, 2009, for France) conclude that higher labor force participation and the narrowing of the gender wage gap will not suffice to produce pension equality in the future and there will be still a need to compensate for child-related career interruptions. Differences in regulations (the extent of noncontributory schemes, the coverage of individual pension schemes, the minimum period of contributions required to establish entitlement, and the retirement age) and in the method of calculating pension benefits may amplify the gendered impact of work life differences (on international differences in pension progressivity see Aggarwal and Goodell, 2013).

12.5.2.1 Coverage and Contributions: The Gendered Impact of Current Reforms

Old-age pension schemes generally include a first minimum level with wide coverage, less demanding conditions, and a benefit level close to that of welfare benefits. Noncontributory schemes are a powerful way to loosen the link between career earnings (and associated contributions) and benefits (Bonnet and Geraci, 2009). They are one of the major tools to fight women's poverty in old age. The Scandinavian countries propose the most generous scheme, with universal pensions for all residents. Most countries offer a kind of safety net but one with conditions and usually income tested; for instance, Canada's "old age security," the United Kingdom's basic state pension (but with very low benefits), the "*minimum contributif*" in France (which is proportional to the duration of the career and is intended to compensate more for low wages than interrupted careers), or the "age pension" in Australia. Women constitute the majority of beneficiaries because of their interrupted careers and low wages.

Beyond these minimum pensions, the retirement income system is generally a mix of publicly administered social insurance (financed by employer and employee contributions) and private pension schemes supplemented by individual savings. The minimum length of employment required for entitlement to a pension may affect women more than men, given their shorter working lives. Pension reforms tend to lengthen the required period, but in some cases the higher labor participation of women has counterbalanced this negative effect, and the gender coverage gap is generally expected to narrow. For instance, the US public scheme (social security) now requires 40 quarters (15 in 1971), but the percentage of women entitled to social security benefits increased from 57% to 65% between 1980 and 1999 (Even and Macpherson, 2004). In most countries, pension reforms tend also to increase the legal age for full retirement benefits; in the United Kingdom, for instance, the state pension age, currently 60 for women and 65 for men, is being gradually equalized to reach 65 for both sexes in 2020. This can be more prejudicial to women than to men because the increase in the length of working life may not be sufficient to keep pace with the increase in the retirement age. OECD pension reforms also tend to align the legal age of retirement of men and women. In the past, the legal age of retirement for men has often been higher than that of women as a way of compensating women for unpaid care work. In 1995, 15 OECD countries applied the same retirement age for men and women and 8 applied different ages; in 2013, only 2 countries (Israel and Switzerland) applied different retirement ages (64 for women, 65 for men) (OECD, 2013). The equalization of the legal age of retirement was favorable for women in higher-paid, full-time jobs, enabling them pursue their careers and to acquire more pension rights (and higher pension benefits), but it can increase income insecurity for women in precarious forms of employment (Luckhaus, 2000).

Another salient trend in pension reforms is to give a larger weight to individual private pension schemes provided through employers, that is, largely OPs, in other words, a switch from first to second pillar pensions (Behrendt, 2000). The link between contributions and the level of benefits is stronger than in public schemes (first pillar), and private pensions are viewed as the main source of the gender pension gap (Bardasi and Jenkins, 2010). This is particularly the case in the United Kingdom, where only 33% of women over 65 years old received a private pension in 1993–1994, compared with 61% of men (Ginn et al., 2001, p. 52). Because the schemes are provided voluntarily by employers, people in part-time employment, precarious jobs, and small organizations are less likely to be covered by OPs. Bardasi and Jenkins (2010) estimated the effects of different variables on the probability of being covered by an OP and found significant correlations between women's OP probabilities and labor status (part-time employment, time out of the labor market); marriage and children have only indirect effects through work history.

In the United States, the correlation between wage rates and pension coverage weakened after enactment of the Employment Retirement Income Security Act in 1974, which limits tax preferences for employer plans for which all workers, including the low paid, are eligible. However, women still tend to have lower rates of coverage because they are more likely than men to work in small firms and in low-paid or part-time jobs (Bajtelsmit, 2006). Similar regulations have been introduced in Europe: two judgments of the European Court of Justice in 1994 stipulated that the exclusion of part-time workers from OP schemes contravenes equal pay laws and may represent indirect discrimination against women.

Another possibility for rectifying gender gaps in pension coverage and entitlements is the policy of matching government contributions for low-income workers (Jefferson, 2009). For instance, Australia has experimented with government cocontributions to personal superannuation accounts. A large proportion of women participate in this scheme, and it has been considered successful in encouraging women to build their private pension savings, although not enough to improve significantly the economic situation of old women (Olsberg, 2006). Moreover, there is doubt about whether it is possible for low-income earners in low-income households to save. It is likely that these savings came from high-income households (where women are in part-time, low-paid jobs), which can take advantage of these profitable cocontributions. In that case, first the policy target is missed, and second, it is not an individualization of private pension savings but a form of pooling savings at the household level (with predictable difficulties in case of divorce).

12.5.2.2 Benefit Calculation Methods

Benefit calculations are identical for men and women, but structural differences in working lives and life expectancies have different consequences on the level of benefits for men and women. Here we examine the consequences of three main characteristics in calculation methods: reference earnings years, defined benefits versus defined contributions, and the choice of annuity tables.

Women are more likely to have low earnings years in their careers, so the longer the reference periods used for benefit calculations, the more they are disadvantaged. For instance, France's 1993 reform gradually increased the number of reference years from 10 to 25 in 2008. Using a microsimulation model, Bonnet et al. (2006) calculated that this reform would reduce the reference wage (which is used to determine the level of benefit) by over 20% for more than 30% of women, compared with 12% of men. Another consequence of this reform is that it increases pension disparities between women.

The pension reforms that shift the focus from the first pillar to the second pillar often are combined with a change from defined benefit to defined contribution schemes (Mackenzie, 2010; Orenstein, 2013). In defined benefit schemes, the benefits are calculated using a fixed formula, and the risk of the amount to be paid is shared by employer and worker. In defined contribution schemes, the payout is dependent on both the amount contributed to an individual account and the performance of the investment vehicle.[96] The expansion of defined contribution schemes is analyzed as a transfer of risk from the employer and the worker to the beneficiary. The majority of pensions provided by private employers in the United States are now based on defined contributions (more than 85% in 2001), and they are expanding in OECD countries. In the United States, they are essentially tax-free saving accounts (i.e., a tax deferred account—retirement plans like the 401(k)), where employees choose whether to participate and how much to save for their retirement; it does not penalize quitters, unlike defined contributions.

Defined contributions offer the advantage of portability, which is appreciated by workers. This can be important for women because of their weaker attachment to the labor market. It is also likely, however, that women will not accumulate sufficient savings over their working lives—they tend to have lower amounts accumulated in their defined contribution plans (because of their lower wages) and therefore lower financial returns. In the event of a separation from their employer, they are more likely to use the lump sum to meet their urgent financial needs rather than roll it over into another tax-qualified savings plan (Bajtelsmit, 2006).

Another growing problem with this shift toward defined contributions is that participants have to be active in their employer's pension plan. Actually, the general observation is that employees are passive and do not change the default rules proposed by employers (see Duflo and Saez, 2003). For women, the negative effect of this passive behavior may be exacerbated by their lower financial literacy compared with men (Van Rooij et al., 2011; see Fonseca et al., 2012, on financial literacy within the couple) and their higher risk aversion.

[96] However, the difference between the two systems has been blurred by the sovereign debt crisis, which led to numerous cuts in pension payments in defined benefits systems.

12.5.2.3 Life Expectancy, Gender, and Pensions

The gender gap in life expectancy also has an impact on individual contributions and benefits in defined contribution schemes, according to the way the benefits are designed. Calculations based on unisex mortality tables have a redistributive effect away from those with lower life expectancy (men) and toward those expected to live longer (women). If the calculations of pensions were based on sex-based actuarial factors, women's pensions would tend to be lower than men's or their contributions would be higher (10–15% higher) (Shilton, 2012). Whether benefits should be calculated on sex-based actuarial factors or unisex tables is an ongoing debate; insurance companies are pushing to be able to discriminate by gender (and their associated mortality tables) and apply gender-based contributions or benefits. The 1978 US Supreme Court ruled in favor of unisex tables, judging that differences in compensation based directly on sex are unlawful as a matter of statutory policy (Luckhaus, 2000) and that contributions and life annuities must be calculated from unisex mortality tables. The underlying rationale is that this ensures equity in the amount the employer pays to the employees, rather than equity in the lifetime value of that benefit to the employee (Jefferson, 2005). A similar debate was held in Europe in the 1990s, and the prohibition of sex-based mortality tables was imposed on insurance companies on the basis of antidiscrimination laws and reaffirmed in 2004 and then 2011 (Shilton, 2012).

12.5.3 Family Matters for Pensions, Too

So far we have examined the gender differences in individual pensions. The big picture is that the individual gender pension gap is due to differences in work history and that the narrowing of labor market participation and wage gaps is not sufficient to reduce gender inequality in pensions "as long as combining paid work and parenthood (or other caring responsibilities) affects women more than men" (Bajtelsmit, 2006). The role of career differences in the gender pension gap has been quantitatively estimated by Bardasi and Jenkins (2010). Using a regression-based decomposition of the average gender private pension income gap in the United Kingdom for people aged 66 and older,[97] they found that differences in returns to personal characteristics account for at least four-fifths of the private pension income gap among recipients. They conclude that this result is probably because of the nature of the jobs taken by these cohorts of workers (part time and of shorter duration). But they also suggest that women in these generations may have chosen not to contribute to private pensions in the expectation that they can rely on their husband's entitlements.

There are strong associations between marital and fertility decisions, pensions coverage, and the level of benefits. Derived entitlements from marital status or family rights are an important component of women's pensions and explain why the gap in living

[97] The data set used is the British Household Panel Survey 1991–2000.

standards during retirement is relatively small compared with the individual gender pension gap. In France, for instance, the median standard of living of elderly women was 10% (single)[98] to 19% (widowed) lower than that of couples in 2009 (Bonnet and Hourriez, 2012).

There is a wide variety of pension entitlements derived from marital status. The US system is particularly generous from this point of view. The social security system includes a spousal retirement benefit, equal to 50% of the pension benefit of the spouse (or former spouse), subject to the condition of at least 10 years of marriage, and a survivor's benefit equal to 100% of the husband's actual benefit. Women's eligibility to these benefits is based on their own earnings history; they receive whichever is higher between their own benefits and the derived pensions. One side effect of this system is that the couple's replacement rate decreases mechanically with the labor market activity of women: one-earner couples have a maximum replacement rate (with a ceiling on monthly earnings) equal to 60% (40% for the husband plus an additional 20% for the spouse); two-earner couples, with individual benefits calculated on their own careers, have a maximum replacement rate by social security of 40% (there is no additional benefit for marital status). Actually, the replacement rate for married households is decreasing across generations from 50% (generations born in 1931–1935) to 45% (generations born in 1948–1953); this decline is being driven by the increase in the frequency of dual-earner couples with their own benefits (Wu et al., 2013). The eligibility condition of 10 years of marriage has a side effect on the duration of marriage: using American social security data, Goda et al. (2007) showed that couples tend to delay divorce decisions from year 9 to year 10.

In Germany and the United Kingdom, there is another type of spousal entitlement whereby pension rights are divided equally between spouses in the event of divorce. The mechanism is intended to remedy the low level of women's own rights (because of unpaid work) and corresponds to a perfect pooling of the pensions rights acquired by the husband and wife during the marriage (Bonnet and Geraci, 2009). Denmark offers an interesting case study in adopting a completely opposite reform in 2006: private pension savings (which form the basis of pension incomes) were no longer regarded as community property in the event of divorce, but as private property. This reform should lead to a substantial decrease in divorced women's pension income because men's pension savings are 30% higher. Amilon (2012) exploited the natural experiment constituted by this change and the period of 7 months, during which couples could divorce without being affected by the reform, and she observed a significant increase in the probability of getting divorced and a rise in women's savings within couples. As in the US case, this tends to prove the sensitivity of family status to the legal pension scheme.

[98] The relatively low gender pension gap for single women is also the result of a composition effect; divorced and unmarried women retirees are, on average, more educated and have accumulated more individual pension rights.

Last, the survivor's pension (defined as a percentage of the deceased spouse's pension) is still used to guarantee the living standard of widows in most OECD countries. It is intended to compensate for the loss of the spouse's pension and to reduce the gap in living standards between couples and widows by compensating widows for the loss of economies of scale within the couple. In a society where couples are stable, this survivor's scheme equalizes the living standards of men and women in retirement (provided retired couples pool income).

However, there is increasing debate about the legitimacy and effectiveness of these spousal rights, especially the survivor's pension, and a decline in these derived benefits. In Sweden, survivor's pensions have been gradually abolished and, more recently, survivor's benefits have been reduced in the Netherlands and Germany (OECD, 2007). Changes in marital history, with the increasing number of divorces and the decline in marriage, explain the need to reexamine the rationale of survivor's schemes (see Bonnet and Hourriez, 2012). There are fewer widows and more spinsters and divorced women in elderly one-person households, so the effectiveness of survivor's pensions in guaranteeing a decent living standard for older women is compromised. As a matter of fact, nowadays single mothers raising children seem to be more at risk of poverty than widows.

Because children and their consequences on mothers' careers are at the core of the gender pension gap, family rights are gaining significance as a way to improve the situation of women in old age, in place of survivor's schemes. Family rights have been extended in various European countries over the past 20 years (Bonnet and Geraci, 2009; D'Addio, 2013). In Italy, Germany, and the United Kingdom, family rights are limited and initially open only to parents who have not worked. France has a more generous system that combines different mechanisms: contribution credits for mothers,[99] old-age insurance for nonworking parents, and pension bonuses for parents of three or more children (this last measure is not aimed specifically at women). The tendency of a shift from marital rights to family rights reflects the decline of marriage and the shortening duration of couples but also recognizes the persistent family-related gender wage gap.

12.6. NONMARKET WORK, THE GENDER DIVISION OF LABOR, AND GENDER INEQUALITY

Together with women's increasing participation in the labor market and commitment to employment, another major and related trend of the second half of the twentieth century was the reduction in the hours of housework, the time women spend at home doing

[99] Contribution credits for mothers are an advantage for women when pension reforms increase the number of quarters that must be worked for entitlement to a full pension, but not when the reforms involve raising the minimum retirement age.

domestic tasks and caring for children or other family members. In the United States, for example, Bianchi et al. (2012) reported a decline in the time dedicated to housework by women aged between 25 and 64, from an average of 30 weekly hours in 1965 to about 16 h in 2010.[100] In this same period, the time spent by men rose from 4.9 to 10 weekly hours. Similar trends have been observed in other Western/industrialized countries (Gershuny and Robinson, 1988; Sayer, 2010; Sullivan and Gershuny, 2001), but with differences in magnitude and timing. The decline in women's share of housework was initially sharper in Anglo-American and Nordic countries than in continental Europe, and since the 1990s the change due to men's participation in housework seems to have been greater in "traditional" countries than in "egalitarian" ones (Geist and Cohen, 2011; Treas and Lui, 2013), as has an increasingly positive social perception about men's participation in housework (Geist and Cohen, 2011).

Despite these trends, cross-country differences remain large and, at the end of the 2010s in OECD countries, the remaining national gender gaps in housework were still impressive, even in the most egalitarian countries: the minimum ratio of women's to men's hours of daily housework, in Sweden, is about 1.3; it stands between 1.5 and 1.8 in many other countries, exceeds 2 in some countries such as Ireland and Italy, and it reaches about 5 in Japan.[101] Beyond housework, women also do more other non-market work such as volunteer work and care for nonhousehold persons (Miranda, 2011). All in all, the share of nonmarket work in women's total hours of work (paid and unpaid) was between 60% (Canada) and 74% (Italy) at the beginning of the 2000s. This means that the major contribution to household production of nonmarket services results from women's unpaid work. The reduction in women's housework time (and the fact that this reduction was far from counterbalanced by the increase in men's time) also results in a much lower total time spent on housework, and in turn in the household production of nonmarket services, than 50 years ago. To sum up: Women still do a higher share of unpaid work than men—albeit of a reduced total amount of unpaid work, and men still do a higher share of paid work than women and have increased their participation in housework. There is very little difference between men's and women's total time of work (Burda et al., 2013; Folbre 2009) but still a significant gender difference in the composition of this total amount of work.

That women carry out the lion's share of nonmarket work is not only well documented in statistics and academic literature; it also seems to be a shared perception, as

[100] Time-use indicators are highly sensitive to definitions and methodology, so the figures provided in this introduction for illustrative purposes are not necessarily comparable between different references. Measurement issues are discussed later in this section.

[101] These figures are computed using the OECD data on time-use available from http://www.oecd.org/gender/data/balancingpaidworkunpaidworkandleisure.htm. The original data, from national time-use surveys, are from various years between 2001 and 2011.

suggested by Burda et al. (2013). In a small ad hoc survey,[102] they asked researchers and students in economics and sociology to give a spontaneous estimate of the difference between men's and women's total work in the United States (including paid work and any activity they related to nonmarket work). About 57% of the respondents estimated that men's total work was at least 5% lower than women's, and 25% thought it was roughly equal (differing by less than 2.5%). Since the question started by recalling that men do more market work than women, this means that the respondents estimated the quantity of women's nonmarket work to be large enough to balance out their lower quantity of market work.

As for the gender gap in total work, Burda et al. (2013) found it close to nonexistent over a sample of 27 countries (the positive difference "in favor" of women is statistically significant but small). They identify this state of things as "iso-work," hence "iso-leisure." However, they wonder about "iso-leisure" when men still work more in the market and their earnings, which are expected to determine their relative power in decision making, are still significantly higher than those of women. "Three logical possibilities present themselves. Men have more power, but are altruistic toward their spouses and toward women generally, and do not take advantage of it. Another is that economists' modeling of the household has been incorrect, and market earnings do not generate power in the household. A final alternative is that earnings do generate power, men are not altruistic, but the average man's utility from his market and home work exceeds that of the average woman's from the same total amount of work" (Burda et al., 2013, p. 258). Another source of puzzlement is that "iso-work" and hence "iso-leisure" somewhat contradict the decline in women's subjective well-being observed by Stevenson and Wolfers (2009), especially its decline relative to men while men's and women's relative opportunities have converged rather than the contrary. This decline in women's perception of their well-being relative to men's is itself consistent with Krueger's "U-index,"[103] showing no real trend over the past 40 years except for "a shift away from activities associated with unpleasant feelings" among men (Krueger, 2007, p. 205).[104] These results, together with the final alternative proposed by Burda et al., suggest that men benefit from a better combination of market and housework than women and perhaps, as we explore later in this section, a better quality of leisure.

Beyond iso-work, which means that men's and women's contributions to productive activities are equal, it is then necessary to consider that the structure of "work" remains

[102] The respondents (far from representative) included 663 faculty labor economists, 255 other economists, 210 researchers and graduate students in sociology, and 533 students in an introductory microeconomics class.

[103] The "U-index" basically measures the share of time spent in activities associated with an "unpleasant state."

[104] Note that Aguiar and Hurst (2008), for the United States, found a decline in women's hours of leisure between 1985 and 2005 and that the trend of increasing leisure time for both women and men since 1965 occurred before 1985.

very unequal, with men allocating a much greater amount of time to market work than women. This may matter a lot: a fundamental difference between market work and nonmarket work is that the former provides incomes in money, whereas the latter does not. The time allocated to unpaid (nonmarket) work relative to paid work, resulting in different outcomes for men and women, therefore has important implications in terms of gender inequality.

This section reviews how the gender division of labor is analyzed today, with a focus on unpaid work. We start by examining the productive dimension of households' nonmarket activity, and the way it appeared at the same time around the 1960s in parallel streams of research on household behavior and on the measurement of economic performance. This leads us to examine some of the methodological difficulties encountered when it comes to measuring and evaluating what happens outside the market and how it may affect estimates of gender inequality. Then we review the effects of nonmarket work—mostly women's time—on the economy and on inequality. The remainder of the section is centered on the analysis of specialization within the household, reviewing how recent empirical research analyzes this very persistent dimension of gender inequality.

12.6.1 Nonmarket Work/Household Production

Standard consumer theory considers households as consumers allocating their resources to consumption and leisure and their members as participating in the labor force to obtain the income they will spend on consumption; standard statistics measure the "economy" on the basis of hours of work, monetary incomes, and the volume of goods and services exchanged on the market. Households and the market interact through labor supply and consumption. These representations changed considerably at the turn of the 1960s with the conjunction of two "revolutions"—a new approach to households as producers and recognition of nonmarket activity as a significant dimension of the economy—at a moment when time-use data started to become available on a regular basis (see Juster and Stafford, 1991). This resulted in notable changes in the conception and analysis of economic performances at macro level, in the analysis of individuals' and households' behavior at a micro level, and in the approach to economic inequality, with particular implications for gender inequality. Whether at macro or micro level, accounting for households' nonmarket production (or its counterpart: individuals' nonmarket work) requires time-use data. This raises specific methodological issues, especially the need to define which nonmarket activities are considered as production or leisure and to find a method of evaluating the time devoted to these activities.

12.6.1.1 Two Conceptual Revolutions

The first revolution, founding the "new home economics," concerns the approach to household behavior with a focus on the productive dimension of this activity. In the 1960s, the interest in home productive activity was not new—early research dates back to the 1930s[105]—but the turning point was Becker's (1965) theory on the allocation of time, in his own words: "a basic theoretical analysis of choice that includes the cost of time on the same footing as the cost of market goods" (p. 494). Instead of the household as a consumer, Becker's theory presents the household acting as a small firm, producing a set of commodities obtained by a combination of time (including time spent on consumption) and market goods; it is these commodities, not goods or time themselves, that provide utility. This approach, which turns consumption into production (and utility maximization into cost minimization), also departs from the traditional approach in terms of a trade-off between work and leisure[106]: substitutions between nonmarket and market time and between time and goods are possible, and "leisure" disappears into a general category of nonmarket time. The agent's constraint, or full income, is then a global resource integrating a goods constraint and a time constraint, and time is allocated depending on its cost relative to the cost of goods in the production of commodities. In multiperson households, the allocation of the household's total time includes an additional dimension, since the time also has to be allocated between the household members, but this allocation remains determined by the same principle of cost minimization in the production of commodities. It is then the relative market efficiency of the household members that determines how they will be assigned to market or nonmarket activity, entailing reallocations if there are changes in the household members' relative market efficiency. Becker's approach was criticized on various points,[107] especially for the lack of a distinction between housework and leisure, "a good approximation of the role the husband plays in the production activity of the household but does gross injustice to the wife (. . .) the wife's allocation of time should therefore be analyzed in terms of a three-way division of work in the market, work at home, and leisure" (Gronau, 1973, p. 634).[108] The new "standard" model of household production finally distinguishes leisure from housework within nonmarket time, following Gronau, basing the allocation of

[105] Zick et al. (2008) provided four references from between 1929 and 1954, including Margaret Reid, who is most often cited for her "Economics of Household Production" (1934).

[106] Mincer (1962) had already underlined the need to take nonmarket activity into account to understand the trend in women's labor supply, which he described as "the most striking phenomena in the history of American labor force [. . .] the continuing secular increase in participation rate of females, particularly of married women, despite the growth in real income" (p. 64); this is actually a trend in total contradiction with the standard theoretical prediction.

[107] See Pollak and Wachter (1975) on the additional assumptions needed for the theory to hold, particularly the need of absence of direct utility from the activity of production itself (or "process benefits"; see Juster et al., 1981). Gronau (1977) also argued that Becker's model was about consumption technology, not household production.

[108] Mincer (1962) had underlined the need to consider "the home-market dichotomy within the world of work" in the analysis of women's labor supply and the family context of decisions regarding labor supply, home production, and leisure.

time within the household on a principle of comparative advantage; we examine later in this section what this implies in terms of the division of labor between men and women.

The second "revolution" happened in the measurement of economic activity, with the recognition of the need to complete conventional measures focused on market goods and market time to take into account households' nonmarket activity. Another limit to conventional aggregates, underlined in gender studies, is that conventional measures neglect a large share of women's contribution to the economy.[109] The revolution took some time, from the 1960s and a growing concern about the shortcomings of the GDP (gross national product [GNP] at the time), to 1993, when the revision of the System of National Accounts introduced the idea of satellite accounts to "enlarge the concept of production to include household production of services for own use" (UN Statistical Commission, 1993). The idea became an explicit recommendation at the UN Conference of Beijing in 1995 (UN, 1995).

The shortcomings of conventional measures of economic activity were the focus of a conference on "The measurement of economic and social performance" held by the National Bureau of Economic Research (NBER) in 1971. In the introduction to the volume of contributions, Moss (1973) speaks of the GNP as "under attack" for neglecting the negative externalities of growth or failing to provide "appropriate measures of economic performance of households and governments." One of the propositions emerging from this conference, and much debated, was to develop accounts for households' nonmarket productive activity based on the measurement and evaluation of households' time. Gronau (1973) proposed an evaluation of "housewives' time." Nordhaus and Tobin (1973) proposed a more encompassing extension, integrating nonmarket time, including leisure, into a "measure of economic welfare." For 1965 in the United States, their lowest estimate for nonmarket activity represented about 42% of the GNP, and their lowest estimate for leisure was about equal to the GNP (Nordhaus and Tobin, 1973, p. 518). Forty years later, the debate is still open, as illustrated by the recent report of the "Commission for the Measurement of Economic Performance and Social Progress" (Stiglitz et al., 2009). Referring to the same shortcomings of the conventional measures as the NBER conference of 1971, the 2009 report recommends "broaden[ing] income measures to non-market activities." The extensions discussed are still intended to provide a better account of well-being, moving a step further "towards a notion of 'full income'" (Stiglitz et al., 2009, p. 126), especially with the valuation of leisure (p. 131) or the development of measures of subjective well-being.[110] The report specifically recommends

[109] This asymmetrical effect of conventional measures had been underlined by Hill (1979a,b), who explained that the types of nonmarket activities typically carried out by women are not taken into account in the "production for own account" conventionally included in the GDP; hence, as he remarks, "the textbook joke" that the GDP goes down when a housekeeper marries her employer because the output is no longer counted, while the output is counted when a mechanic replaces the family car's engine.

[110] See Fleurbaey (2009) and Decancq et al. (2014), Chapter 2, in this volume .

taking into account the "services that households 'deliver to themselves' such as child care, cooking or parent's education services to children" (Stiglitz et al., 2009, Recommendation 5, p. 14). The 2009 report is, however, not a repetition of the NBER conference of 1971, and at least two points go further: First, it underlines the need to address gender differences in time use and the issue of the distribution of unpaid domestic work within families—a view that was absent from the debates of 1971. Second, it includes an investigation into subjective well-being, a strand of research that has attracted increasing attention since the 1990s and goes beyond GDP not by developing extended indicators, but by changing the units of measurement to account for individual perceptions and affects—this is the idea behind "well-being accounts" or "national time-use accounting" (Kahneman et al., 2004; Krueger et al., 2009).

12.6.1.2 Measurement and Valuation Issues

The two revolutions differ in their scope and implications, but they are connected by the same "subject": housework or its output, that is, the services that households provide to themselves. But nonmarket work is not captured by the usual indicators of participation in the labor force. Being unpaid, it does not appear in individuals' or households' money income; its output is consumed but there is no market transaction, and hence it does not appear in households' expenditure. No market hours, no wage, no money income, no price and a product in unobservable quantity: economists and statisticians, then, are deprived of their favorite measurement kit. Basically, the problem is to define which nonmarket activities should be considered as productive, to measure the time spent on them,[111] and to convert it into money. These questions have been debated for decades, with a focus on the evaluation of "productive" time outside the market—our focus here.[112]

[111] A general difficulty in measuring time is how to obtain detailed and reliable information. Time-use survey practice is beyond the scope of this chapter, but discussions can be found in Juster and Stafford (1991), Hamermesh et al. (2005), and in international guidelines (Eurostat, 2008; UN, 2005). One important issue is the mode of data collection (diary versus other methods); there is a long-standing consensus among experts that the former is preferable (see a recent comparison in Kan, 2008). Sampling is another issue: It is possible to obtain aggregate estimates by surveying only one person per household, but this does not provide useful data for the analysis of intrahousehold allocation of time (see Apps, 2003). Finally, time-use estimates, and hence gender gaps in time use, are very sensitive to the type of household (individuals living alone or in a family, presence of children), age group, and occupation, as well as whether gender gaps are measured between men and women on average or between men and women within the households.

[112] Measurement methodology has been especially debated in experimental accounts of household production (see, e.g., Chadeau, 1992; Landefeld and McCulla, 2000). See a review of various options by Eurostat (2003) and a discussion of these methodological options by Poissonnier and Roy (2013). Note that households' satellite accounts include measures of nontime inputs (contribution of capital services). We also concentrate on "domestic" work rather than "unpaid work" (see Miranda, 2011), which includes nonmarket activity outside the household (volunteer work or services provided to persons outside the household).

12.6.1.2.1 Perimeter

Which nonmarket activities are "productive"? There is no clear theoretical definition, and practice depends on whether measurement is implemented to analyze individuals' (households') behavior or in a national accounting perspective. In the first perspective, the focus is on individuals' (households') activity, especially in terms of choices of time allocation; in the second it is on the product of this activity.

When focusing on individuals' activity, the problem is to distinguish leisure from productive activity within nonmarket time. A first approach is that of the household production model—in Gronau's (1977) revisited version,[113] since Becker's did not distinguish leisure within nonmarket time: "An intuitive distinction (. . .) is that work at home (like work in the market) is something one would rather have somebody else do for one (if the cost were low enough), while it would be almost impossible to enjoy leisure through such a surrogate" (p. 1104). In this view, nonmarket work is different from leisure because, unlike leisure, it does not provide utility directly but only through its outcome (provided the absence of process benefits is assumed, as underlined by Pollak and Wachter [1975]). Other approaches to the contrary, in welfare economics or time-use research, emphasize the "emotional" content of various activities, that is, whether the time spent on it is enjoyed or found pleasant, precisely process preference. Stiglitz et al. (2009, p. 135) evoke this research as a promising way to be able to distinguish between leisure and production in household activities.[114] Folbre (2004) proposes a more encompassing approach to nonmarket work, arguing that the valuation of time does not fully account for the positive externalities—not only within but also outside the household—that can result from it.

The accounting perspective does not rely on a notion of preference or pleasantness but adopts a reference to the output and the market; nonmarket work is generally identified by the fact that it provides services that could have been purchased in the market. According to the literature (e.g., Ironmonger, 1996, 2000; Juster and Stafford, 1991; Zick et al., 2008), Reid (1934) is considered to have been the first to propose what is referred to as "the third party" (or person) criterion: "If an activity is of such character that it might

[113] Gronau's (1977) objective was to explain why changes in wages did not affect husbands and wives work/leisure trade-off in the same way, suggesting that other factors operate: "A wage increase may result not merely in a shift from work at home to work in the market but also in a reduction of leisure—employed women having less leisure than the not employed (. . .). A natural question is to what extent will the labor supply function of married women resemble that of their husbands once they reach similar labor force participation rates. Right now any answer to this question should be regarded as sheer speculation, since so much depends on changes in role differentiation and reallocation of work at home within the family" (p. 1117).

[114] It is debatable whether this is operational as an aggregate because it would involve moving work/leisure boundaries between individuals. It may also be difficult to combine with material aspects of well-being, for example, to account for the double benefit of enjoying the time spent on market work *and* being paid for it.

be delegated to a paid worker, then that activity is deemed productive" (cited by Ironmonger, 2000, p. 6). This leads to the identification of a set of nonmarket activities broadly including preparing meals, cleaning and laundering, caring for children or other household members, managing the home, repairing the home, gardening and caring or pets, and related shopping and travel.

The various approaches are not necessarily incompatible, and none is completely straightforward.[115] Economic or time-use studies sometimes restrict the perimeter of nonmarket work, for instance, by only keeping routine or "core" domestic work, excluding gardening and pet care (which may have a leisure character) or shopping (its direct relation with the productive activity can be difficult to check). Childcare is often treated separately (one reason being that time cannot be spent on childcare of one's own children in childless households) and sometimes is refined into "active" and "passive" childcare (or primary and secondary, depending on whether another activity is performed simultaneously; see Allard et al., 2007).

Of course, variations of the perimeter (often related to the perspective of the analysis or the issue addressed, as well as to the accuracy of the information on time use) have a significant impact on the "quantity" of nonmarket productive time measured and on the estimate of men's and women's contributions to this quantity or gender gaps in time use. Using the French Time Use survey from 2010 on the population aged 11 and older, Poissonnier and Roy (2013) compared three perimeters. Starting with a restrictive definition limited to core activities (cooking, washing, cleaning, active care of children and dependent adults and related travel time, household management), they obtained a daily average of 2 h 7 min; an intermediate definition, adding shopping, home repair, gardening, and playing with children—or "productive leisure"—results in almost one more hour (3 h 4 min), and an extensive definition, adding driving oneself and pet care, adds about 50 min more (3 h 53 min). In these daily averages, the share of women's time decreases significantly when the perimeter is extended, with values of 72%, 64%, and 60%, respectively. Using the American Time Use survey from 2010 on the population aged from 25 to 64 years, Bianchi et al. (2012) obtained a similar pattern in the ratio of women's to men's time, which falls from 2.9 when the perimeter is restricted to "core" housework to 1.6 for total housework.

These differences reflect the fact that, on average, men and women do not do the same sorts of housework. In short, "women cook, clean and care while men build and repair" says the OECD (2011b, p. 22). Beyond averages, the time spent on housework and care by men and women varies with the composition of the household. Of course, the presence of children has an important effect on the total time spent on care,

[115] Chadeau (1992, p. 89) discusses "borderline cases" that the third party criterion does not solve, for instance, care activities, which are considered productive when performed for a sick or disabled person but not for oneself, although it is the same service.

but it is not the only factor; in the United States, married men and married women spend more time on total housework than the average, and even more on core housework, and women spend more than men; the gender ratio (women's time to men's time) for this perimeter is 3.4 versus 2.9, on average (Bianchi et al., 2012). Parents naturally spend more time on childcare, and mothers spend more than fathers (the ratio is 1.9), but they also both spend more time on core housework than the average married men and women—again mothers spend more time than fathers, signaling that the presence of children tends to reinforce the unequal gender division of nonmarket work (Bianchi et al. 2012). The difference is also pronounced in the detailed activities related to childcare: 60% of mothers' childcare time is devoted to physical care and supervision versus only 45% of fathers' childcare time; conversely, 41% of fathers' time is spent on educational and recreational childcare compared with 27% of the mothers' time (Bianchi et al., 2012). In addition, men's and women's "specialization" within nonmarket work may be weighted differently because of the routine versus occasional nature of the core or other housework (Sayer, 2010) or because of a difference in the emotional content of the activities—a direction explored by Stevenson and Wolfers (2009). "Women's" housework may also be more intensive than men's if several activities can be carried out simultaneously (e.g., cooking and doing the dishes and looking after the children) or interrupt their leisure time (Bittman and Wajcman, 2000; Sayer et al., 2004). Such differences do not count in an accounting perspective but may count in the perspective of measuring and analyzing gender inequality—hence the interest of going beyond the simple measurement of quantities of time.

12.6.1.2.2 Valuation

Once the relevant activities have been defined and the time spent on them measured, the next step is to obtain a value of the time or product. There are two main approaches to this valuation. The "replacement cost" method consists of estimating the price the household would have paid to obtain the same service on the market (or the money saved if the service is produced at home instead of purchased on the market). The "opportunity-cost" method estimates the earnings the person (or the household) would have received if the time devoted to housework had instead been spent working in the market. As in the perimeter definition problem, there is no generally agreed best method, and either might appear more or less appropriate depending on its application.

The opportunity-cost method is generally used in the analysis of intrahousehold organization and decision making and inequality within the household; it is consistent with the methods used for the valuation of leisure. But it is problematic in the accounting perspective because it results in the same output having different values depending on who produced it (Eurostat, 2003; Landefeld and McCulla, 2000). Using the opportunity cost can also be problematic when the producer of the service does not work in the market. This problem has been discussed at length. Gronau (1973) tested two estimates under the alternative assumptions that women who do not work in the market are more efficient in the home sector or less efficient in the market sector, resulting in a price of time respectively higher or lower than the potential wage rate. Bonke (1992) used the individual wage rate in the case of working men and women and the reservation wage for those who do not work in the market. Another direction of discussion is that the opportunity-cost method assumes that the individual would work for pay instead of doing housework, but this approach neglects the fact that market work is often not offered by the hour. The choice may then be between taking a job or not, or between a full-time or part-time job, and then giving up more hours of market work than the number of hours actually spent on nonmarket work. We cannot exclude the possibility that at the individual level, housework might cost more in terms of foregone leisure time than in terms of foregone earnings; while this does not change the valuation at the household level, it maybe not be so in terms of individual well-being.

The replacement cost method has the advantage of being consistent with the "third person criterion": the cost of what is available on the market does not depend on the wage of the person buying it (while of course, this wage is relevant in the choice of doing versus purchasing). Eurostat (2003) distinguished two main approaches to the replacement cost: the "specialist" cost (using the wages of specialized wage workers—a cook in a restaurant, a nurse at a daycare center, etc.—or the cost of specialized services at home—e.g., a private nurse, cleaning person, plumber[116]) and the "generalist" wage (sometimes referred to as the "housekeeper's" wage). In addition, there is the option of using the minimum wage (when there is one). None of these solutions is fully satisfying (see also Landefeld and McCulla, 2000), first, because in the household, several tasks can be carried out simultaneously, while the corresponding service would require several specialized workers; second, some tasks cannot be carried out by "generalist" workers (e.g., fixing the roof). Another possible drawback of the replacement cost method is the need to find market substitutes for all the services the household provides to itself.[117] It is possible that some of these services are specific to the point of not having market equivalents; more generally, market substitutes can be of poorer (or better) quality than homemade products (not to mention the emotional dimension of some domestic or care tasks; see Folbre, 2004).

As with variations of the perimeter, choosing one or another method of valuation can result in wide differences. The opportunity cost always yields a higher value for household production of nonmarket services than any other method. Over a sample of

[116] This may result in higher estimates than in the case of wage workers because it includes other costs than wages.

[117] Gronau (1980, p. 414) also objected that using market prices can be irrelevant when the household does not consume the market service because it assigns to the household production prices that are rejected by the household.

25 countries, it is at least twice as high as the replacement cost (housekeeper rate) in 7 countries, between 1.5 and 2 times as high in 10 countries, and between 1.1 and 1.5 as high in the 7 remaining countries (Ahmad and Koh, 2011; link to data in OECD, 2011b). Landefeld et al. (2009, p. 219) estimated the value of nonmarket household services in the United States using various options. For 2004, the lowest value is obtained using the minimum wage; compared with this value, the "housekeeper's" wage gives an estimate 54% higher, the "specialist" wage gives an estimate twice as high, and the opportunity cost gives an estimate five times higher.[118] At the individual (household) level, the difference between the two values (market price vs. opportunity cost) illustrates the possibility that the increase in money earnings from working for pay instead of producing nonmarket services could be higher than the decrease in the value of home production.

12.6.2 Taking Households' Production and Productive Time into Account: What Does It Change?

As mentioned earlier, one of the aims of measuring household production—responding to long-lasting debates on the limitations of conventional aggregates measuring economic activity,[119] income, or consumption—was to provide a full account of economic and social progress and an accurate basis for the measurement of inequality. At a micro level, the conventional approach, which considers only money income (or expenditure), does not provide a satisfying estimate of the household well-being either because it neglects the actual access to goods and services, which is enhanced by household production. Given the average contribution of men and women to the production of nonmarket services, taking household production into account also allows one to take full stock of the impact of women's work, especially its distribution between paid and unpaid work, on economic well-being. This impact is rather different depending on the perspective: unpaid work reduces inequality among households but, because it is mostly women's work, it contributes to gender inequality.

12.6.2.1 Household Production and the Markets

At the macro level, it has been shown that household production of nonmarket services represents significant shares of GDP: in 2008, from 15% in Canada to 26% in the United Kingdom, with a replacement costs valuation, and from 40% to 68%, respectively, with the opportunity-cost method (Ahmad and Koh, 2011). The shares of household production tend to be negatively linked to participation in paid work in general (see Miranda, 2011), if only because total time is constrained. Very basically, more time in paid work is

[118] The differences between the various options of valuation were less pronounced in 1985, showing that the choice of an option can also affect comparisons over time.

[119] Gronau and Hamermesh (2006) also remark that the interaction between time and goods and its variations (in other words, the demand for more or less goods-intensive commodities) may have additional implications in the relative demand for traded goods at the international level.

associated with more market goods, and less time in paid work (hence more time in nonmarket production) is associated with more household production—even more if one concentrates on women's participation. So as long as there are significant cross-country differences in the distribution of time between work and leisure and, within work, between market and nonmarket work, taking household production into account tends to have an "equalizing" effect in cross-country comparisons of economic well-being. For instance, the disposable income per head is higher in the United States than in France by 21% points, but only by 16 points when corrected for household production (in 2005; figures from Stiglitz et al., 2009, p. 131). Comparing Finland, France, and the United States between 1995 and 2006, Stiglitz et al. (2009) also remarked that the annual growth rates of households' extended income, that is, corrected for household production, are significantly smaller than those obtained without this correction—in other words, not correcting for household production results in overestimating growth in well-being. The same was found by Landefeld et al. (2009) over a longer period (1965–2004) when comparing the growth rate of US GDP and an extended measure including nonmarket household services; they interpreted this as reflecting women's entry into the labor market, which resulted in an increased growth rate of market production, but neglecting its counterpart in terms of reduced household production overestimates the increase in production.

The relationship between household production (i.e., the unpaid time spent on housework and care) and market products depends first on how the population is distributed by household type (especially singles versus families) and on the degree of substitutability between household products and market products. It also depends on whether men's and women's nonmarket time are perfect substitutes in the household utility function, and it has been found that this is not the case. This is one implication of the tests rejecting the unitary model and a result of other precise tests (see, e.g., Browning and Meghir, 1991, who find that the assumption of substitutability is strongly rejected). Previous analysis of labor supply also had shown that men trade market work for leisure, whereas women trade market work for both housework and leisure (e.g., Killingsworth and Heckman, 1986). This may have changed—as we have seen, men's housework time tends to increase (albeit slowly), but all in all, this points to women's allocation of time as the main driver of changes in both labor supply and the relative shares of household products and market products in the consumption of commodities. In turn, changes in the demand for market products have employment effects in the sectors producing market substitutes for home production.

This is the question addressed with the "marketization hypothesis": Freeman and Schettkat (2005) linked the larger amount of time worked in the United States than in the European Union with a more pronounced departure from household production in the United States and the shifts it induces in terms of both "freeing" time and increasing the demand for market substitutes—hence the employment in the sectors producing these substitutes. In turn, the jobs created in these sectors contribute to the increase in women's employment (although this is primarily low-skill employment; see Section 12.3). Freeman and Shettkat showed that the difference between the United States and European countries is mostly attributable to women's allocation of time to paid work versus housework; as the authors remark, "substitution of male for female time in housework is modest" (p. 37), and, in the case of men, the trade-off is essentially between paid work and leisure. A striking difference is the time spent by women on childcare, which is much lower in the United States, where the proportions of children younger than 3 years old in formal daycare is more than twice higher than the average in their sample of EU countries. Other studies (Del Boca, 2002; Jaumotte, 2003; Wrohlich, 2004) showed that childcare availability could be more important to the trade-off between market work and household production than its cost (together with other arrangements that help to reconcile work and family responsibilities; see Section 12.3).

There is also an ongoing debate in time-use research on the interaction between women's entry into paid employment, that is, a departure from housework, and expenditure on time-saving domestic appliances (the "mechanization" hypothesis) or reliance on market substitutes (the "buying out" or "outsourcing" hypothesis). New technologies, assumed to increase the productivity of housework, should allow women to devote more time to paid employment. But various studies have found that, contrary to the intuition, either the time spent on housework remained constant, or "the more technology, the more time spent" (Gershuny and Robinson, 1988, quoting themselves, 1972). Likewise, using the Australian Time-Use Survey from 1997, which also provides information on household equipment, Bittman et al. (2004) did not find a significant effect of ownership of appliances on housework or on its division between genders (it even seems to go with more pronounced gender specialization within housework). As for time-saving services (food away from home, childcare, paid domestic services, laundry services, etc.), Bellante and Foster (1984) found a positive association between women's participation in paid work and buying out, but no relation with the number of hours worked, and most of all a strong relation with the family income. With time use data but a small sample of dual-earner couples, Killewald (2011) found that the relation between the use of market substitutes and housework time (excluding childcare) is, at best, very small (e.g., an increase of 1% point in expenditure on market substitutes was associated with a reduction of 1 min per week spent on cooking and cleaning). All in all, studies provide inconclusive results (see a review by De Ruijter et al., 2005)—probably partly because of differences in data quality and methodologies.

12.6.2.2 Household Production and Inequality Between Households

Across households, studies have regularly found home production to be less unequally distributed than money income (Aslaksen and Koren, 1996; Bonke, 1992; Frazis and Stewart, 2011). This is consistent with theoretical prediction: provided there is no process preference, high opportunity-cost individuals will turn to more goods-intensive consumption and, conversely, low opportunity-cost individuals will turn to more time-intensive commodities and therefore spend more time on household production. Taking into account nonmarket production then tends to result in lower estimates of economic inequality than those obtained with conventional measures of income.

The "equalizing" power of nonmarket work might have somewhat declined over time, as has been suggested over the last decade or so.[120] Returning to the main link between women's allocation of time to market or nonmarket work, it seems obvious that the increase in women's participation in market work goes together with a decrease in household production. However, this may not be completely obvious; one must assume that nonmarket time was not initially totally allocated to leisure (time-use evidence confirms this) or that the increase in market time was not just added to an unchanged amount of nonmarket work (time-use evidence confirms this, too, contrary to what the notion of the "second shift" suggested[121]). Zick et al. (2008) found that, in the United States, the correction due to household production was larger in the 1970s than in the early 2000s. They argued that the shift in women's time from unpaid to paid work has two opposing effects: on the one hand, women's entry into the labor market has slowed the increase in money income inequality; on the other, high-wage women have reduced their housework time more than low-wage women, hence creating an increased inequality across households, which is partly limited by the increase, even if small, in time spent on housework by men in high-income couples. These changes have been reinforced by marital homogamy (high-wage men marrying high-wage women), contributing to inequality between couple households. At the same time, changes in family composition, especially the increase in single-parent households (most often headed by women), have reinforced inequality (see a discussion by Kollmeyer, 2013). These results are consistent with those obtained by Gottschalk and Mayer (2002); comparing the 1970s and 1980s, they also showed an equalizing effect of household production for both periods, albeit larger at the bottom of the distribution.

This highlights the fact that the size of the equalizing effect also depends on the level of inequality in the distribution of households' money income, which is itself influenced

[120] Most comparisons of the distribution of income versus extended income do not take into account household size, or they only consider couples. But some provide estimates for households of different composition (e.g., Aslaksen and Koren, 1996); the size of the "equalizing" effect is larger than that, on average, for multiperson households and households with children. This adds a demographic dimension to the trend of the declining equalizing effect of household production. Note that including household production in a measure of extended income may impact the equivalence scale. This has been examined mainly through the cost of children (see Apps and Rees, 2002; Craig and Bittman, 2008); Couprie and Ferrant (forthcoming) estimate economies of scale in the time allocation of the household.

[121] *The Second Shift* is the title of the very influential book by Hochschild and Machung (1989). Drawing attention to the addition of paid work to domestic work in the 1970s, it nevertheless somewhat overestimated the gap between women's and men's total work.

by women's earnings (a counterpart of the decline in women's unpaid work). This influence is essentially studied through counterfactual experiments: what if women had no earnings, or if their earnings were not unequal? Comparing couples' earnings in 1979 and 1989 in the United States, Cancian et al. (1998) found for both years that including wives' earnings reduced the inequality in family money incomes, that is, inequality would have been larger if wives had no earnings. However, the level of inequality would have been lower than observed in the absence of inequality in wives' earnings.[122] Comparing income inequality in 16 countries in the early 2000s, Harkness (2013) found that first, despite large differences in households' employment structure and the incidence of part-time work among women, the contribution of women's earnings to household income is quite similar across countries—with the exception of the United States, where this contribution is the largest at close to 30% of gross income. Three counterfactuals are examined. If no women worked, household earnings inequality would increase in all countries, most of all in the United States. If all women worked, there would be a decline in earnings inequality. The third counterfactual consists in closing the gender gap (all else unchanged); here, the results are mitigated, with a decrease in inequality in some countries and an increase in others—especially in Nordic countries—but it would decrease among couples. The analysis also underlines the varying effect of women's earnings in different types of households, especially between one- or two-earner couples and between couples and one-person households, especially single women. Beyond the general result, a case study of Italy (Del Boca and Pasqua, 2003) also shows that the "correcting effect" of women's earnings is highly dependent on the level of married women's participation in paid work. Working on a counterfactual "if no women worked," Del Boca and Pasqua found that income inequality would have been much higher than observed (in 1977, 1989, and 1998) in northern Italy, where wives' employment is higher, whereas it would have been comparable in southern Italy, where fewer married women work. They underlined the fact that southern Italy lacks many of the factors facilitating women's access to employment, especially opportunities for part-time work and childcare facilities, resulting in large inequalities in women's earnings. It would be especially interesting to combine these results with an analysis of the correcting effect of nonmarket work because Italy is one of the OECD countries where women's participation in the labor market is the lowest and the amount of housework is the highest (see OECD, 2011b). Unfortunately, no such study yet exists.

[122] Men's earnings also play a role. Juhn and Murphy (1997) explained that the wives of high-wage men significantly increased their participation in employment in the 1970s and 1980s compared with the 1960s, while at the same time low-wage men's earnings decreased sharply and their wives' participation in the labor market slowed (on the contrary, it was faster in the 1960s when their husbands' earnings were also growing quickly).

Nevertheless, despite its probable decline, the equalizing effect of household production remains quite large. Folbre et al. (2013), who compared couples' market earnings and a measure of extended earnings including nonmarket work (valued at the minimum wage[123]) over nine countries, showed that the Gini coefficients with the measure of extended earnings are reduced, in most countries, by about one-third compared with those for the earnings. There is some variation between countries, but the rankings remain fairly stable—except for the United States,[124] where the equalizing effect is the lowest because of the combined effects of a relatively low share of nonmarket work in total work and a low minimum wage (the lowest of the nine countries relative to average earnings).[125]

12.6.2.3 Housework and Gender Inequality

We turn now to the influence of housework on gender inequality. The issue is not only its unequal distribution between men and women per se, but also its consequences on other socioeconomic outcomes. One, pointed out by Becker (1985), is that it may affect women's earnings: "(. . .) the earnings of men and women would not be equal even if their participation [in paid work] were equal" (p. S35). This is not because of discrimination in the workplace but the fact that housework and childcare demand so much energy from women that little is left for market activity—a hypothesis that would save human capital theory from the "embarrassment," in Becker's words ("failure" in the terms of England, 1982), of not explaining gender earnings differentials. Another is that the "pressure of time" may affect women, who have to combine paid work and family responsibilities, more than men; the "dual burden" or "second shift" would then result in a gender gap in leisure (Hochschild and Machung, 1989; Bryant and Zick, 1996). While this gap is almost closed (on average) in terms of overall quantity, recent work points at a gender gap in the quality of leisure time. Ultimately, both approaches suggest that housework and family responsibilities influence the capacity to focus on and/or benefit from what one is doing when not doing housework. If this is so, it may affect gender inequality in ways that conventional measures of individual living standards or well-being do not capture and may contribute to women's disadvantage in economic and social outcomes.

[123] Using the minimum wage is debatable: it may result in under- or overestimation of the value of time, depending on its level; as mentioned earlier, no method of valuation has yet been proved to be better than another, but, because the purpose is to compare households, it would be interesting to test whether other methods (e.g., opportunity costs) might change the rankings.

[124] Finland, France, Germany, Italy, Poland, Spain, Sweden, the United Kingdom, and the United States.

[125] This and Harkness' (2013) results suggest that there can be large differences between what is observed in the United States and what is observed in the rest of the world. Given the large share of US-based evidence in the literature, it is an invitation to be cautious about whether the trends observed in the United States are representative of more general trends, not in terms of a "big picture" but in terms of the magnitude and timing of changes.

12.6.2.3.1 Housework and Wages

According to Becker (1985), women's family responsibilities affect their potential for effort at work because these tasks drain their "energy." The more energy spent on child-care and housework (which he describes as being highly demanding in energy), the less left for the job. Hence, a choice of less demanding and, accordingly, less well-paid jobs or those with lower career opportunities, as analyzed in the literature on the family pay gap since the mid-1990s (see Section 12.3). However, strictly speaking, the family pay gap is an effect of the family status, especially motherhood, not an effect of housework (even if the time spent on housework is not independent from the family status); moreover, the focus is on differences between women. Then arises the question of whether there is an effect of housework that is distinct from the effect of motherhood or family status and of its impact. Following Becker, such an effect should be negative on women's wages; what is slightly unclear, however, is whether this effect operates indirectly, by influencing the choice of hours and type of job (work effort), or directly, by the performance at work (effort on the job) once the occupation, hours worked, and so on have been controlled for. As emphasized by Bielby and Bielby (1988), the notion of "energy" refers not to the quantity of time or the occupation but to the intensity of physical or mental effort required on the job. As for men, the logical complement to Becker's proposition is that married men (or, more generally, men living in couples) should be able to allocate more energy to their work because they do not suffer the same energy-draining home activity as women and they benefit from their wives' housework.

Testing the assumption of an intrinsic impact of housework on wages is difficult because one must be able to measure the expenditure of energy on both housework and paid work, or to control for the jobs' actual requirements (e.g., concentration, precision, availability), and to combine this with accurate information on housework. It also poses various estimation problems, particularly because of the potential endogeneity of wages (wages have regularly been found to influence negatively the time spent on housework) and unobserved heterogeneity (a detailed review of the econometric problems is provided by Maani and Cruickshank, 2010).

A first question concerns the impact of housework on the expenditure of "energy" at work; this was the question asked by Bielby and Bielby (1988). Their test used an index of work effort based on self-reported information on the demands of the job (physical and mental) and whether the actual effort made goes beyond what is required. Their index was designed to avoid a potential gender bias, for example, exhausted women overreporting the job's requirements. However, the authors argue, the sociopsychology literature suggests that women tend rather to understate their performance, so that if there is bias, it is more likely to be in the other direction. Their main result is that, controlling for education, occupation, age and family composition, working time and pay, family responsibilities (responsibility for child, childcare hours, domestic work hours), and whether the spouse works, women did not allocate less effort to work than men. Married women or mothers allocated less effort than single women, but this reduction in effort left them at the same level as men. In the case of men, on the contrary, the hours of housework seem to reduce slightly their effort at work, suggesting a trade-off that did not occur in the case of women: Women's efforts in paid work do not depend on their hours of housework, but men's do. However, the result for men could be a case of reverse causality, that is, men's effort in housework might depend on whether they work in more or less demanding jobs. Another limitation, more likely to affect the results for women, is that the estimation (by a sample of men and women working at least 20 h per week) did not account for possible selection into paid work; it could well be that the most exhausted women do not work in the market. This study by Bielby and Bielby remains, to date, the only study directly tackling the issue of the allocation of effort.

Hersch's (1991) study of the United States was one of the first to examine the effect of housework on wages. Controlling for various job attributes (hardship, autonomy, stress), she found substantial differences between men and women in the frequency of these attributes but no significant difference in the returns to them. Housework (self-reported) had a negative effect on women's wages (not men's), but only as long as the job attributes were not controlled for. Moreover, the negative effect was found only for housework on work days. Since then, almost all empirical studies have found a negative relation between housework and women's wages (see a detailed review of empirical work from the 1990s to the 2000s in Maani and Cruickshank, 2010). Hersch and Stratton (2002), also studying the United States, also found that including housework in a decomposition of the gender wage gap substantially increases the explained part of the gap. However, the size of the effects found is highly sensitive to the estimation strategy; the review of results by Maani and Cruickshank (2010) shows that using ordinary least squares (OLS) yields the highest effects, using instruments and two-stage estimation seriously reduces the effect (but satisfactory instruments are difficult to find), and using fixed effects reduces the effect even more. Moreover, some results differ substantially from the dominant evidence, finding no effect of housework on wages. McLennan (2000) found no significant effect once the endogeneity of time spent on housework was controlled for (using the ideal number of children in her instruments—information that is seldom available); but this result could also be because her results were obtained from a sample of younger men and women (mean ages of 34 and 33, respectively) than in comparable US studies. Hirsch and Konietzko (2013), studying German data, found "no impact at all" of housework on wages; contrary to the majority of other studies relying on self-reported housework time, they used detailed diary data matched with register data on earnings.

A number of studies from the 2000s also examined whether the type of housework tasks, especially their frequency, flexibility, and timing, has different effects on wages. Noonan (2001) and Hersch and Stratton (2002) found that frequent tasks or inflexible tasks—that is, those that cannot be postponed (and are mostly performed by women)—are the main drivers of the negative impact of housework on wages and of its stronger effect for women than for men. When the tasks are not disaggregated, Noonan (2001) did not find any gender difference in the negative impact. Bonke et al. (2005), using Danish data, showed that the timing and flexibility of housework have an impact on wages; using quantile regressions to estimate the wage functions they also found a surprising positive relation between housework and women's wages in the 90th quantile of the wage distribution and a negative effect on men's wages—contrary to their result for the rest of the distribution. The impact of housework also seems to be different depending on the type of working time; for instance, Bryan and Sevilla-Sanz (2011), using British data, found a negative impact of housework on wages except for women working part-time, and Bonke et al. (2005) found a stronger effect for workers whose working time schedule is not flexible.

As for the relation between housework and men's wages, it is less clear, and, when an effect is found, it is weaker than for women. Moreover, Hersch and Stratton (2002), for instance, found that men's hours of housework do not influence the "marriage premium" (see also Section 12.3); but Gray (1997) and Chun and Lee (2001), both using US data, found a negative relation between the hours of wives' paid work and men's marriage premium. At first view, these results seem to be conflicting; on closer examination, they suggest that men's wages are influenced by the amount of housework done by their wives rather than by themselves, which would fit with the view that women's specialization in housework boosts men's productivity in market work. This would also be consistent with the difference in the composition of housework performed by men and women: daily, repetitive tasks (most often women's) can result in more constraints than occasional tasks (most often men's) or tasks that can be postponed. But there is also unsupportive evidence of such an effect. Lincoln (2008) did not find a relation between men's marriage wage premium and the division of housework. More precisely, neither the hours of market work nor the hours of housework performed by wives significantly affect the husbands' wages, but Lincoln found a small positive effect of the hours of housework performed by husbands on wives' wages.

Finally, is there an effect of housework on wages? Empirical results suggest a negative impact of housework on women's wages rather than no impact, and Maani and Cruickshank, 2010 argued that the bulk of results cannot be suspected of being driven only by endogeneity bias and unobserved heterogeneity. Housework would then be a useful dimension to take into account in explaining the gender wage gap. But beyond the issue of econometric problems, the main limitation of existing studies is that most of them, including those published in the 2000s, are based on US data from the 1980s or early 1990s. Evidence from more recent data, and from other countries, would help to clarify the issue. It would also be particularly interesting to analyze more systematically the relation between men's housework and wages, now that they seem to participate more in housework, as well as the impact of housework among younger men and women. The fact that, as mentioned above, no impact of housework was found by Bryan and Sevilla-Sanz (2010) or Hirsch and Konietzko (2013) in countries (Germany and the United Kingdom) where the incidence of part-time work in women's employment is especially high, also call for further investigation.

12.6.2.3.2 Market Work, Housework, Leisure, and Well-Being: Is ISO-Work Equality?

Most of the literature on the effects of housework on time use has stressed that paid work on top of domestic work particularly increases women's burden; this was the message in *The Second Shift* (Hochschild and Machung, 1989), attracting attention to the fact that women's additional hours of paid work were not counterbalanced by an equal reduction in their hours of domestic work, even though the figure they provided of women's weekly hours of total work (paid plus unpaid) exceeding men's by about 15 h on average is highly unlikely, as underlined by Sayer et al. (2009). Whether performed in the market or in the home, work has the same effect of reducing the amount of free time remaining, but some analyses suggest that domestic work (housework and childcare) may affect this remaining time, and thereby men's and women's subjective well-being, in ways that do not appear in the amount of leisure. A first issue is that of the reality of unequal or equal total work. Using the most recent data available from the OECD,[126] it seems that many countries are actually close to iso-work (Table 12.4); in the 16 countries examined, only 5 had a gender gap in total work of more than 15 min per day (in Sweden, the Netherlands, and Denmark, total work seems to be greater among men than among women). However, iso-work covers highly different gender gaps in paid and unpaid work. Japan, where the gap in total work is the lowest, is also the country where the gender gaps in paid work and domestic work are the highest. On the contrary, large gender gaps in paid work do not necessarily result in large gaps in total work, as illustrated by the case of Italy versus Ireland. The next point that can be made from this sample of countries is that the gender gap in leisure is negative in all the countries examined, that is, women have less leisure than men (by only a very small amount in the case of Japan and by an amount smaller than the gap in total work in the Netherlands). Burda et al. (2006) also found lower amounts of leisure for women using data from the early 2000s and, distinguishing between weekdays and weekends, much higher gender gaps in leisure on weekends.

First, this suggests that, contrary to what was implicitly acknowledged earlier in this section (and with all necessary caution, because the computation is based on second-hand data), iso-work does not necessarily mean iso-leisure; in other words, on average, women derive less leisure from each hour of total work than men (see Figure 12.7). Second, the difference between weekdays and weekends (which cannot be examined in the data from the OECD) also suggests that comparing total amounts of time does not reveal all the dimensions of gender differences in time use.

[126] Data available from OECD (2012), updated March 2014.

Table 12.4 Gender gaps (women − men) in work and leisure in selected Organisation of Economic Co-operation and Development countries (hours and minutes per day)

Country (survey year)	Market work	Domestic work	Total work	Leisure	Burda et al. (2006)[a] Year	Weekday	Weekend
Sweden (2010)	−0h58	0h27	**−0h30**	**−0h41**			
Netherlands (2005–2006)	−2h11	1h48	**−0h23**	**−0h17**	2000	−0h02	−1h00
Denmark (2001)	−1h15	1h02	**−0h13**	**−0h20**			
Japan (2011)	−3h49	3h48	**−0h01**	**−0h04**			
Spain (2009–2010)	−1h35	1h38	**0h03**	**−0h37**			
Australia (2006)	−2h13	2h16	**0h03**	**−0h28**			
Germany (2001–2002)	−1h43	1h47	**0h04**	**−0h25**	2001–2002	−0h14	−1h07
Canada (2010)	−1h18	1h30	**0h12**	**−0h42**			
Austria (2008–2009)	−2h02	2h14	**0h12**	**−0h36**			
Belgium (2005)	−1h25	1h38	**0h13**	**−0h43**			
United Kingdom (2005)	−1h42	1h56	**0h14**	**−0h42**			
United States (2010)	−1h08	1h23	**0h15**	**−0h46**	2003	−0h05	−1h01
Ireland (2005)	−2h24	2h49	**0h15**	**−0h51**			
Finland (2009–2010)	−0h41	1h13	**0h32**	**−0h50**			
France (2009–2010)	−0h59	1h31	**0h32**	**−0h48**			
Italy (2008–2009)	−1h45	2h56	**1h01**	**−1h03**	2002–2003	−0h54	−1h50

Population aged 15–64 years except in Australia (≥15 years) and Sweden (25–64 years).
Paid work includes work in all jobs, job search, and related travels; domestic work includes routine housework, shopping, care for household members, and related travels. The data do not allow the exclusion of travel related to study or to unpaid work other than domestic.
[a]Ages 20–74.
Source: authors' calculations based on Organisation of Economic Co-operation and Development data, http://www.oecd.org/gender/data/balancingpaidworkunpaidworkandleisure.htm.

However, these results "on average" may not precisely capture the notion of "double burden" or "second shift" because this notion refers not to average men and women but to couples and especially to parents—hence only a fraction of the populations examined above. One might expect the gender gaps in work and in leisure in these cases to be more pronounced than on average; in the United States, married women (with or without children) perform more core housework than women on average, and married men perform more noncore housework than men on average (Bianchi et al., 2012). Both married fathers and married mothers carry out not only more childcare than average married men and women but also more core housework (women more than men). The presence of children also has been found to affect mothers' allocation of time to paid and domestic work more than fathers'. Maume (2006), using US data, or Sayer et al. (2009), using Australian and US data from the late 1990s, show that men do not adjust to their partners'

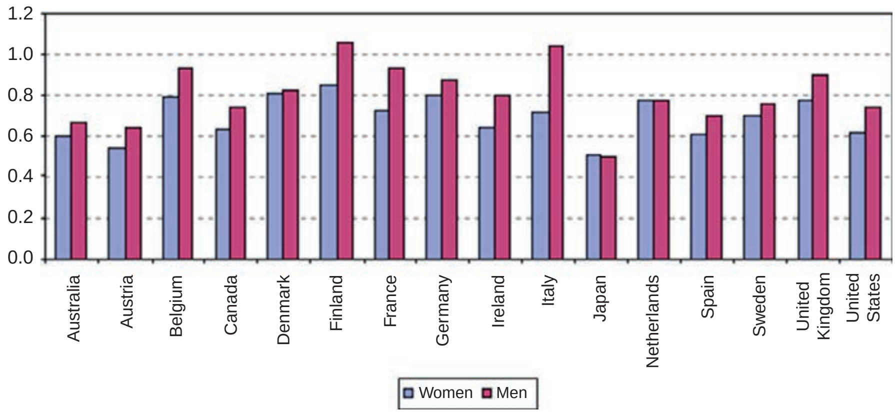


Figure 12.7 Ratio of leisure to total work by gender in selected Organisation of Economic Co-operation and Development countries.

hours of employment; Gershuny et al. (2005), however, using panel data from the United Kingdom, Germany, and the United States, suggest that men gradually adjust their participation in housework to changes in their spouse's employment, even though this does not result in a complete reallocation of housework. Finally, various studies have found that parenthood reinforces a traditional division of labor between spouses (Sayer, 2005 on the United States; Craig, 2006 on Australia; Hallberg and Klevmarken, 2003 on Sweden; Régnier-Loilier, 2009 on France; or Craig and Mullan, 2010 on the United States, Australia, Denmark, France, and Italy). Comparing the United States and three European countries, Anxo et al. (2011) observed that having children causes the "greatest revolution in the time that individuals spend in unpaid work." Change may be underway: Dribe and Stanfors (2009) find a change in the effect of children between the 1990s and the 2000s in Sweden. In a comparison of 20 countries from 1965 to 2003, Hook (2006) found significant changes in men's housework and care that were positively associated with women's participation in market work, and cross-country variations were influenced by the share of women working part-time, the length of parental leave (longer parental leave discourages men's participation), and whether parental leave is available to men. Yet, for the time being, all studies found that women's use of time is more affected by their family status than men's.

In addition to gender differences in the amount of time spent on housework and the possible impact on a gender gap in the amount of leisure, several studies underline that women's leisure time could also be "less leisurely" than men's (Bittman and Wajcman, 2000), or more fragmented and "contaminated" by other tasks (Mattingly and Blanchi, 2003). Multitasking is also more prevalent among women than men and substantially increases the time spent on housework (Craig 2007; Offer and Schneider, 2011); while it makes it possible to squeeze more activities into a same span of time, it may contribute to feelings of time stress. All these analyses point to the pervasive nature of "family responsibilities"—still, today, more incumbent on mothers than fathers—which fits quite well with Becker's view of energy-draining activity, but more in the sense that women (mothers) do not benefit fully from their leisure time than that it prevents them from participating in market work.

Analyzing the quality of leisure requires detailed data not only on the primary use of time but also on secondary activities to detect simultaneous activities and whether leisure is actually a time free of constraints; it also requires reliable measures that are available only from detailed individual data (see a comparison of measures, and subsequent invitation to caution, by Lee and Waite, 2005). Bittman and Wajcman (2000) resorted to notions of "pure" and "adult" leisure, that is, excluding a secondary nonleisure activity, and measures of the length and number of leisure episodes and whether children are present during these episodes. Using Australian diary data from 1992, they found that men have an advantage in all of their indicators of leisure quality, most of it related to the presence of children, and the fact that parents specialize in different types of care: mothers in primary care and fathers in more leisurely care (see Miranda, 2011). Mattingly and Blanchi (2003) obtained similar results using US diary data from 1999. They also observed that men's leisure time with children is more often spent in the presence of others, whereas women's is more often a time spent alone with their children, and that free time is more correlated with reduced feelings of time pressure among men than among women. Focusing on multitasking, Offer and Schneider (2011), using US data collected using beepers,[127] argued that it may allow spouses in dual-earner families to spend more time and more enjoyable time together with their children but, again, that mothers are more likely than fathers to multitask in the presence of children. Their data do not accurately represent all families in the United States (and constitute a rather small sample), but they have the advantage of including measures of emotional state over the day being surveyed. Controlling for work and family characteristics, they found that subjective well-being is positively affected by multitasking in the presence of one's spouse and that feelings of work–family conflict among mothers are reduced when multitasking in the presence of children.

Finally, does housework affect perceived well-being differently for men and women? Very few data sets provide reliable measures of both time use and information that allow one to compute any measure of subjective well-being (feelings of stress, perceived time pressure, or dissatisfaction). Consequently, empirical results are obtained using rather unreliable measures of either the dependent or the key independent variables, making it difficult to answer the question of the impact of housework. Among the small number

[127] This method of data collection (the experience sampling method) consists in providing participants with programmed beepers that randomly ask the respondents to indicate what they are doing, where, and with whom, and in this case how they were feeling, several times per day over 7 days.

of studies addressing the issue, two use a measure of time stress (how often the respondent feels rushed or pressed for time), and obtain quite contrary results despite very close methodologies: Hamermesh and Lee (2007) studied couples with at least one partner in the labor market (they mention the same results for two-earner couples) in Australia, Germany, the United States, and Korea, using data that provide self-reported (not diary) hours of market work and housework. They find that more hours of domestic work only increases time stress for women, but the effect is only significant in Australia and Germany; the same was found for the effect of the partner's market work. One's own market work increases time stress for both men and women. Time stress also seems to be correlated between the partners.

Using the same dependent variable (time stress) on Danish data for two-earner couples, Bonke and Gerstoft (2007) found no significant effect on time stress of either one's own and partner's hours of housework or of one's own and partner's hours of market work. However, hours of domestic work have a negative coefficient for both women and men, which Bonke and Gerstoft interpreted as a sign that housework acts as a destressor. They used an additional control variable, which they call the "rush hour," measured as the length of time between market work and housework at the end of each day (excluding commuting time); a short length of this time is interpreted as a sign of time pressure. When this variable is introduced, hours of housework become significant for women; they analyzed this rather unexpected effect as possibly reflecting women's "preferences" for housework because of their "more family-oriented perspective" (p. 58) and speculated that it gives support to the idea that the work–family conflict affects women more than men. The nature of this time between market work and housework, excluding commuting time, seems at first rather unclear; it is not free time—in this case, the results would only imply that less rushed women (or women who have more leisure) feel less rushed by housework; the only other possibility is that it corresponds to "physiological time" (e.g., a shower or a snack).[128]

All the other studies examined (MacDonald et al., 2005, on Canadian data, using a measure of time stress and a measure of satisfaction with work–family balance; Boye, 2009, comparing 25 European countries and using a composite measure of psychological well-being; Mencarini and Sironi, 2012, comparing 26 European countries and using a measure of satisfaction with life) found a negative effect of housework on women's measure of subjective well-being. Overall, with the exception of Bonke and Gerstoft (2007), the studies examined regularly obtain a significant coefficient (of the expected sign according to the dependent variable) on the hours of housework. The results also strongly suggest that while hours of paid work increase time stress (again, except Bonke and Gerstoft, 2007), they are positively associated with psychological well-being or life

[128] This analysis could be an invitation to question the classification of activities between leisure and physiological time.

satisfaction for women, although none of these effects are significant for men. Another interesting result is that once the hours of housework have been controlled for, the presence of children has no significant effect (MacDonald et al., 2005; Mencarini and Sironi, 2012) or a positive effect (Bonke and Gerstoft, 2007, in their specification with the "rush hour"). However, one weakness of these studies is either their dependent variable (time stress is only a partial approach to subjective well-being or satisfaction with life) or their measures of housework (the last three studies reviewed do not use information from diaries but derive their measure of housework from information given by one respondent on the hours of housework in the household and his or her share of it). Here again, more research using appropriate data on both the dependent and independent variables would provide more precise results.

12.6.3 Within the Household: The Persistent Gender Division of Labor

So far, gender inequality in the time allocated to paid and unpaid work has just been taken as given. At the end of this section, we turn to a last but no less important question: Why do women do more unpaid work than men? We start with an overview of the economic and sociological approaches to the division of labor within the household and then examine how empirical research has, since the 1990s, also focused on the question of why, when so much has changed in women's market activity, so little seems to have changed in men's domestic activity. A last part addresses the issue of the "quiet" but unfinished revolution using comparative research.

12.6.3.1 An Overview of the Theoretical Background: Approaches to the Division of Labor Within the Household in Economics and Other Social Sciences

There is no shortage of theoretical tools to analyze the division of labor: economic theory uses rationality, other social sciences focus on power relations, and both use social norms. Whether grounded in economic theory or other social sciences, they are not mutually exclusive; beyond the difference of perspective there is a strikingly large number of common features—and predictions.

12.6.3.1.1 The Economic Approach: Specialization and the Allocation of Time to Paid and Unpaid Work

In economics, the basic framework was introduced by Becker's theory of household production (see Section 12.6.1), in which the household members (henceforth we speak indifferently of spouses or partners) allocate their time according to their relative market productivity to maximize the household's full income. How time is allocated is not a question of gender (Becker, 1985) but of efficiency—a choice of technology (Lundberg, 2008; Pollak, 2013; Pollak and Wachter, 1975). Each spouse specializes in the activity (market work or housework) in which they have a comparative advantage. Moreover, Becker (1965) argued, no more than one member of the household should work in the market. This is a rather extreme view of specialization, which may have been an idea of the American family back in the early 1960s, but is not very operational nowadays (in particular, it is very difficult to interpret part-time work in this conceptualization: Is it specialization or nonspecialization? Is it expected to be efficient or inefficient?). Comparative advantage is in principle not "gender," but Becker's analysis is somewhat unclear when it comes to women's "intrinsic" advantage, their "biological commitment to the production and feeding of children," completed with a rational consumer's behavior—that is, expecting returns to one's investment and adding a productivity argument: it is quite easy to produce additional children while nursing those already born (Becker, 1991, pp. 21–23).

Becker's approach to the family has been criticized at length (see Section 12.2), and we will not develop the argument further. However, one particularly problematic aspect is that, depending on the point of view, spouses are sometimes assumed to be perfect substitutes (if they are equally productive in the market) and sometimes not (because of the "natural" advantage women have in childcare). A critical point in the perspective of gender (in)equality[129] is that, in principle, how time is allocated to market work or housework does not depend on whether the person is a man or a woman, and that the link between the allocation of time and efficiency is conceived at the level of the household as a whole. From an individual point of view, is it rational to specialize in housework? It depends essentially on whether income pooling and sharing is such that each household member gets their fair share and whether they stay together—or move to another household. Marriage as a formal commitment is then an important condition of stability; the distribution of income within the household is another. Grossbard-Shechtman (1974), in line with Becker (1973), proposed a view of marriage as an exchange of housework based on the comparative advantages of the partners; Grossbard-Shechtman (2003) introduced a notion of quasi-wage as a means to share the gains of specialization; at the equilibrium, the husband's supply of market work equals the wife's demand, which depends on the opportunity-cost of her time in housework.

In nonunitary models, earnings (precisely, wages as prices; see Pollak, 2005) and relative earnings are among the determinants of the spouses' bargaining power or threat point, depending on the model (see Section 12.2); power relations are made explicit instead of technology. Introducing household production does not fundamentally change the underlying principle (Apps and Rees, 1997; Browning and Chiappori, 1998; Chiappori, 1997), but unpaid work can be analyzed as a source of bargaining power exchanged for consumption/resources. This is difficult to test, however, because there are no data on individual consumption/expenditure. An empirical alternative is to

[129] Another is that of the necessity of specialization for the outcome to be efficient. See a discussion by Pollak (2013).

consider leisure as the outcome (e.g., Beblo and Robledo, 2008; Browning and Gørtz, 2012; Couprie, 2007; Datta Gupta and Stratton, 2010).

To sum up, the main prediction of the economic approach is that, one way or another, wages and housework should be negatively correlated, either because an increase in a spouse's earnings results in a reallocation of time from unpaid to paid work (time-consuming commodities become more costly and household production decreases) or because an increase in a spouse's earnings results in an increase in his or her relative power. This relation from wages to housework, together with the possible relation from housework to wages (see Section 12.6.2.3.1) looks like a chicken-and-the-egg question: If domestic work impairs labor market outcomes and poor labor market outcomes imply more housework, this is rather depressing from the perspective of gender equality. The issue has been addressed in a few theoretical models (e.g., Albanesi and Olivetti, 2009; Attanasio et al., 2008; Cigno, 2008; Francois, 1998; Ishida, 2003), all of which conclude that policy interventions can break the vicious circle.

12.6.3.1.2 Sociological Approaches to Housework Time: Dependence, Power, and Gender

A major difference between the economic approach and other social science approaches is their scope; while the economic approach provides a framework to analyze the allocation of time between paid and unpaid work, these approaches seek to explain the time of housework. Two main theoretical approaches are generally identified in the literature: "time availability" and "resource exchange" theory.

The approach in terms of time availability (Coverman, 1985; Hiller, 1984; Shelton, 1992) considers that housework time depends on the time left available to each spouse after deduction of their time of paid work. It therefore takes paid work to be quasi-exogenous. The underlying model is a unitary model, that is, the spouses are motivated by the interest of the family. One weakness is that the time spent in paid work may depend on factors that contribute to the division of housework, especially children.

Resource exchange theory basically posits that the organization of the household depends on the relative power of the spouses, which is itself determined by the resources they bring to the marriage (Coleman, 1988; Kamo, 1988; Safilios-Rothschild, 1970, 1976).[130] This approach allows for an unbalanced interdependence; in other words, A's utility may depend on B's resources more than B's utility depends on A's resources. Moreover, this unbalanced relationship allows B to impose decisions on A. Sticking to the perspective of the distribution of housework, resource exchange means that the spouse who brings in more money income (or other resources) is expected to provide

[130] The focus here is on the allocation of time to housework. Resource exchange theory is not limited to this issue and has many other applications in sociological approaches to family organization and marital power, in feminist theory, and in gender inequality in general, which are beyond the scope of this chapter.

economic support and to obtain, in exchange, more housework from the other spouse—this is the basis of the approach in terms of "dependence" (Brines, 1994) and a formulation of the "male breadwinner" model.

While the scope of the resource exchange approach is not the same as that of the economic approach, resources exchange, rational specialization, and the collective model all link the domestic sphere and the market in one way or another. The underlying principle of interdependence brings the collective approach and the resource exchange approach especially close together, and both rest on two basic principles: first, that the allocation of time (housework) is related to the partners' bargaining power, and second, that a form of income (earnings, wages, possibly assets) plays an important role in determining the spouses' respective bargaining power. As for the explanation of housework time, the three approaches predict a direct or indirect interaction between the spouses' relative resources, more precisely that the amount of domestic work carried out by one spouse should decrease if his or her wage/earnings/bargaining power increases and vice versa. This prediction has not systematically been obtained in empirical work, leading research in a third direction: the influence of social norms.

12.6.3.1.3 Social Norms: Doing Gender and Identity

Since the widely cited article by Akerlof and Kranton (2000), using social norms to explain economic behavior has become increasingly frequent (Davis, 2006). The concept of identity helps to explain behavior that may at first seem to deviate from economic rationality in the sense that individuals make choices resulting in apparently poorer outcomes than if they had made other possible (rational) choices (see Section 12.3). From the perspective of housework, their identity model allows for an asymmetrical relation between the share of paid work and the share of unpaid work carried out by spouses if social norms prescribe that a man should earn more than his wife—and should not do a woman's work (i.e., housework). In this framework, the cost in terms of "male identity" of earning less than one's wife (a deviation from the norm) helps to explain why lower-earner men do less housework than their wives (and conversely why higher-earning women nevertheless do more housework than their husbands).

The same idea, albeit not in terms of utility function, was developed earlier by West and Zimmerman (1987) in an influential article entitled "Doing Gender"—almost a trademark in time-use research on wages and housework time (see the survey by Coltrane, 2000). West and Zimmerman highlighted the social dimension of gender roles and used the example of housework as typical of an activity socially perceived as "women's work"; hence, for a woman, doing the housework is conforming to the social norm of what a woman should do—whatever her wage might be (South and Spitze, 1994, provide one of the first quantitative applications). "Doing Gender" posits that gender (understood as socially defined gender roles) can have an effect even in the absence of unequal resources in the same way as the identity model suggests that individuals can choose actions that seem irrational from the point of view of their economic outcome. It is only recently that both articles have started to be cited in the same articles—another case of parallel strands of literature addressing the same questions but not communicating with each other.

A last approach related to social norms, known as "gender ideology," links intrahousehold arrangements to attitudes about work and family roles.[131] The distribution of housework resulting from these attitudes defines various categories ranging from "traditional" to "egalitarian" and predicts essentially a positive relation between traditional gender ideology (i.e., the male breadwinner model) and women's housework time (and, conversely, a negative relation with men's housework time).

12.6.3.2 An Overview of Empirical Results

As we have seen, the theoretical toolkit available for empirical work offers many options. However, this multitude of options can be a problem because of the similarity of many of their underlying ideas. Sociological approaches in terms of relative resources, gender ideology and time availability seem to find support in research on the US (Davis and Wills, 2013); but explanations in terms of economic dependence seem challenged in cross-country comparisons (Davis and Greenstein, 2004). All posit that spouses' allocation of time to housework should vary according to their relative market productivity (measured by their relative wages) or their relative power (measured by their relative earnings or bargaining power) following a unique pattern: more housework should indicate less power, more power should result in less housework, and partners' behaviors should be symmetrical. But beyond the broad pattern, things can be less straightforward. First, from an economic point of view, a change in the relative wage rates or relative earnings of the partners has not only an effect in terms of power but also a substitution effect and an income effect; second, the relation from wages/earnings to housework can be mitigated by gender or identity effects.

12.6.3.2.1 Some Methodological Issues

Beyond broad similarities, empirical studies differ considerably in their methodologies (see Table 12.5), including the dependent variable, which most often is hours of housework but sometimes the share of housework. Housework is defined in more or less inclusive ways; however, most studies exclude childcare. The explanatory variable of main interest can be the partners' relative earnings or a measure of the woman's contribution to the couple's earnings, most often as a measure of economic dependence (following the

[131] A related approach is developed in Hakim's (1998, 2000) "preference theory", but this mostly aims to explain women's participation in paid work. In particular, she suggests that preferences may explain why, despite different social structures and social policies across European countries, the patterns of women's participation in the labor market are so similar. She provides a classification of women in three groups (home centered, adaptive [the majority], and work centered), which has led to much debate (e.g., McRae, 2003).

Table 12.5 Methodological differences in investigations of the determinants of housework time

Author (year), country	Data	Hours	Population	Perimeter	Measure of hours	Main independent variable
Brines (1994), United States	CS 1983–1985	Self-report	Couples Age ≥18 in 1983	No childcare	Amount weekly	Dependence[a] (current and long-term)
Greenstein (2000), United States	CS 1987–1988	Self-report	Couples Age <65	No childcare	Amount weekly	Dependence
Schneider (2011), United States	CS 2003–2007	Diary	Married men and women Age 18–65	No childcare	Amount daily	Dependence
Gupta (2007), United States	CS 1992–1994	Self-report	Married women Age 18–65	No childcare	Amount weekly	Annual earnings Relative earnings
Gupta and Ash (2008), United States	CS 1992–1994	Self-report	Couples Age 18–65	No childcare	Amount weekly	Annual earnings Relative earnings Wage
Killewald and Gough (2010), United States	Panel 1976–2003	Self-report	Married men and women Age <60 Employed FT	No childcare	Amount weekly	Annual earnings Earnings share
Bittman et al. (2003), Australia	CS 1992	Diary	Couples Age <55	No childcare, shopping excluded	Amount daily converted to weekly	Dependence
Connelly and Kimmel (2009), United States	CS 2003–2004	Diary	Married parents of children aged <13	Nonpaid activities	Amount daily on weekdays and weekends	Wage Relative wage
Baxter and Hewitt (2013), Australia	Panel 2001	Self-report	Women living in a couple	No childcare	Amount weekly	Absolute earnings Relative earnings
Halleröd (2005), Sweden	CS 1998	Self-report	Two-earner couples with children	—	Shares of domestic work and childcare	Dependence Relative wage
Evertsson and Nermo (2007), Sweden	Longitudinal 1991 and 2000	Self-report	Couples Age 20–56 in 1991	No childcare	Amount weekly	Dependence

Alvarez and Miles (2003), Spain	CS 1991	Self-report	Two-earner couples	No childcare	Amount weekly	Relative earnings
Sevilla-Sanz et al. (2010), Spain	CS 2002–2003	Diary	Two-earner married couples, both working full-time Age 20–65	—	Amount daily	Relative wage
Kalenkoski et al (2009), United Kingdom	CS 2000–2001	Diary	Couples with children aged <18	Only childcare	Amount daily	Wage Relative wage
Bloemen and Stancanelli (2014), France	CS 1998	Diary	Couples with children		Amount daily	Wage
Evertsson and Nermo (2004), United States, Sweden	CS 1974–1999	Self-report	Couples Age 18–65	No childcare	Amount weekly	Dependence
Gershuny et al. (2005), Germany, United Kingdom, United States	CS and longitudinal	Self-report	Couples Born <1932; 1932–1951; 1951 or after	No childcare	Wife's weekly share	Employment status
Esping-Andersen et al. (2013), Denmark, United Kingdom, Spain	CS Early 2000s	Diary	Couples Age 25–60	—	Share daily	Relative wage

CS, cross-section.
[a]Sorensen and McLanahan's (1987) measure.

measure proposed by Sorensen and MacLanahan, 1987),[132] or absolute earnings, or relative wages. The sample of couples can be more or less restricted according to age or employment status; sometimes, it is not a sample of couples but a sample of married men and women. Most data sets are cross-sectional or use a cross section of panel data.

As always with time-use data, measurement is a serious issue (see earlier in this section), especially considering the difference between self-reported hours and diary data and the definitions of housework (e.g., Coltrane, 2000; Sullivan, 2011). Measurement error is also very likely because few (if any) data sets provide reliable information on both time use and earnings or wages. In particular, many results are based on self-reported housework time,[133] and it has regularly been found that this results in under- or over-reporting. Finally, issues of unobserved heterogeneity are also very likely to arise.[134] A last point to be clarified is that most empirical work describes variations across couples, not changes within couples[135]; the effect of relative power or wages is inferred on the basis of the observed distribution at one point in time.

12.6.3.2.2 The Case of "Doing Gender"

Most of the existing empirical work can be related to Brines' (1994) finding of a nonlinear relation between partners' relative resources and their hours of housework. Basically, wives' hours of housework are negatively related to their relative resources (i.e., the larger their contribution to the couple's earnings, the lower their hours of housework), and husbands' hours of housework are not really affected, but only up to the point where their wives' relative resources become larger than their own, that is, the point where they become dependent on their wife. Past this point, husbands do less housework than their wives. Brines analyzed this asymmetry as "gender display." Since then, a large strand of research has focused on what happens when a wife has greater resources than her husband, confronting the "dependence"/"relative resources" approach and the "doing gender"/"gender display" approach. The type of pattern observed by Brines has been found quite regularly for the United States (Bertrand et al., 2013; Greenstein, 2000; Evertsson and Nermo, 2004; Schneider, 2011), in Spain (Alvarez and Miles, 2003; Sevilla-Sanz et al., 2010), in Australia (Baxter and Hewitt, 2013; Bittman et al., 2003), and in Sweden (Halleröd, 2005). The interpretation is that a deviance-neutralization

[132] (Woman's earnings − Man's earnings)/(Woman's earnings + Man's earnings).

[133] Sometimes even reported by only one spouse for both partners.

[134] For instance, the marital status. Various studies show that married or cohabitant couples do not adopt the same financial organization (see Section 12.2); cohabiting and married couples also have been found to differ in the division of labor (Baxter, 2005; Davis et al., 2007; Domínguez-Folgueras, 2012; South and Spitze, 1994), as have couples married after cohabitation (Batalova and Cohen, 2002; Baxter et al., 2010). This suggests selection effects (see Barg and Beblo, 2012) that are most often not accounted for.

[135] This would mean being able to observe whether, within couples, changes in one partner's resources result in changes in the distribution of housework time.

process occurs, with the wife exaggerating her "woman's work" or the husband reducing his contribution—"inflection points" have been found from 50% to 75%—and husbands, or wives, contribute more or less in neutralizing the deviance. For the United States, Brines and Greenstein found that it is rather husbands who contribute, but Schneider found that it is mostly wives; for Australia, it seems to be wives.

The broad picture is that husbands' and wives' housework time is believed to result from the combined effects of relative resources and gender; as long as wives remain the lower earners, relative resources operate, but more strongly for women than for men (men's housework time is generally found to be rather flat). Past the point of deviance, gender talks, and men, or women, act to neutralize the gender deviance. But it seems that this result is not systematic. For instance, Evertsson and Nermo (2004) obtained the standard result for the United States in 1991 but not in 1999. For Sweden, they did not find that higher-earner wives did more housework, but among two-earner couples with children, Halleröd (2005) found that dependent husbands tended to do less housework; moreover, a significant share of husbands with lower market productivity than their wife nevertheless had higher earnings: instead of doing more housework, they worked longer hours.

The scope of a "gender-deviance neutralization effect" has been much questioned since the mid-2000s on other grounds than (sparse) cross-country evidence, starting with serious methodological drawbacks highlighted by Sullivan (2011). England (2011), drawing on later research on the United States and Australia, argued that focusing on the point where the wife out-earns the husband resulted in "missing the big picture." Several limitations have been underlined.[136] First, the "gender-deviance neutralization effect" is at best very small, both in magnitude and because of the small share of couples above the inflection point (for instance, only 4% in Australia, according to Baxter and Hewitt, 2013). Second, the neglect of absolute earnings leads to confusion between higher-earner wives and high-earning women (Gupta, 2007). Third, related to the neglect of absolute earnings, the relation between earnings and housework is not the same for low-earning wives and high-earning wives, partly because high-earning wives have already reduced their hours of housework and outsourced many basic tasks (Killewald and Gough, 2010).

12.6.3.2.3 Changing Patterns?

Contrary to the broad picture conveyed by the previous results, with husbands either not responding to changes in the economic status of their wives or husbands or wives resisting their threatened gender identities through the performance of housework, other research based on longitudinal data or using wages instead of earnings provides a more nuanced picture.

[136] Including methodological drawbacks (measurement first), but also mis-specification, which can account for spurious results (see Sullivan, 2011).

Comparing Denmark, the United Kingdom, and Spain, Esping-Andersen et al. (2013) found a significant relation (with the expected sign) between spouses' relative wages and men's paid and unpaid work in the United Kingdom and Spain but not in Denmark. The division of labor across couples is also more heterogeneous in Spain and the United Kingdom than in Denmark. Rather than lending support to an interpretation in terms of "doing gender," they suggested that this is an indication of more or less stable and widespread equilibrium related to different stages of the "ongoing revolution of women's roles." The ongoing revolution involves adaptations in men's roles, too. Gershuny et al. (2005), using longitudinal data, analyzed whether spouses' allocation of time to housework responds to a change in employment status; they found that, when the wife starts a full-time job, both wives and husbands (but more slowly) adjust their time of housework, resulting in an increase in the housework carried out by the husband. Also using longitudinal data, Evertsson and Nermo (2007) found that changes in spouses' relative resources between 1991 and 2000 had a small negative effect on wives' share of housework, mainly because of an increase in husbands' time spent on housework, resulting in men doing a larger share of the (smaller) total amount of housework.

Studying French parents' simultaneous time adjustments to differences in wages, Bloemen and Stancanelli (2014) found that for fathers and mothers, own market and nonmarket hours are related in the expected way to their own wages. They also found a negative relation between the father's wage and the mother's market hours, but no significant relation between the father's wage and the mother's nonmarket time. Mothers' hours of paid work then seem to depend on the partners' relative market productivity, but their hours of housework depend only on their own market productivity, and their parental time does not seem affected by their wage. Father's parental time, however, seems responsive to their wife's wage; in other words, mothers cannot increase their housework/childcare hours further, but fathers can, suggesting that a continuing trend of reduction of the gender wage gap could result in a reduction in the gender housework and childcare gap. Focusing on childcare in the United States, Connelly and Kimmel (2009) found that both spouses seem to adjust their childcare time to their partner's paid hours or relative wage on weekdays—but not on weekend days—suggesting substitution effects. In the United Kingdom, Kalenkoski et al. (2009) do not find such an effect: men's time use seems rather unresponsive to their wife's and even their own wage, but there is a significant positive association between their own wage and their time of primary childcare on weekend days.

Taken together, these results suggest that another social norm, under which housework or childcare is not a marker of gender identity, could be emerging. This is a possible interpretation of results from research into the divergence between spouses' reports of their own and partner's hours of housework. As mentioned earlier, self-reported housework time is particularly likely to be biased by overreporting, and it has regularly been found that both men and women are prone to overreporting (Coltrane, 2000; Kamo, 2000). Comparing self and partner's reports in 35 countries, Geist (2010) showed that there are considerable cross-country differences in the disparities in reports of both men's and women's housework time, including for the division of housework in specific tasks. She found that discrepancies are stronger for men's housework hours than for women's: men self-report more hours of housework than their partners say that they do, whereas the difference is much less pronounced in the case of women. Kamo (2000) argued that such discrepancies are related to frustrations and social desirability, which brings us back to norms. Geist (2010) suggested that men overestimate their own contributions to "fit better the model of a supportive husband, even though they may not do much work at home"; however, she also found that discrepancies are greater in countries where the overall time of housework is higher. Treas and Tai (2012) also found that men emphasize shared household management but that shared decision making is more prevalent in more egalitarian countries, where women's participation in the labor market is also higher. These findings fit rather well with the analysis by Gershuny et al. (2005) or Esping-Andersen et al. (2013) in terms of lagged adaptation and changing equilibrium.

12.6.3.2.4 Changing Contexts

While most of the empirical research on intrahousehold organization has focused on individual-level (or household-level) explanations, as soon as a comparative perspective is adopted it is obvious that contexts matter. There are actually few reasons to think either that what happens within households is completely unrelated to a broader environment or that "gender" talks universally. Many researchers share the idea that comparative perspectives—across comparable countries—is central to a better understanding of what shapes the patterns of labor division within households. Some even argue that it is more important than individual-level factors (Treas and Lui, 2013). Whether it is more or less important can be debated; however, virtually all comparative research observes that patterns of the division of labor in couples differ across countries, generally according to welfare regimes and the impact of public policies on the work–family balance (Aassve et al., 2014; de Henau et al., 2013; Fuwa and Cohen, 2007; Geist, 2005; Gornick and Meyers, 2003; Hook, 2010; Ruppanner, 2010; Sayer, 2010). Whether referring to Esping-Andersen's (1990) typology of welfare states or another type of clustering, all found that greater gender equality, in terms of labor market participation and access to individual rights not mediated by the family structure, is associated with more equal/less unequal sharing of housework within couples.

In terms of policies, the main focus is on those that facilitate mothers' participation in employment, essentially the provision of public childcare,[137] parental leave policies, and

[137] High-cost childcare raises two different issues. The first involves inequality between high-income and low-income families because the former but not the latter can sidestep the issue. The second is a lingering question about who pays for childcare (the same question can be raised more generally about outsourcing); in terms of opportunity costs, the outcome of women's labor supply is implicitly weighted as their wage net of childcare costs. To what extent do women actually pay for the arrangements that allow them to do paid work?

policies aiming to increase fathers' participation in childcare (see Section 12.3).[138] On average, childcare is not the largest component of housework time, but having children, in addition to the immediate changes induced (see Anxo et al., 2011), has probably the most pervasive effects in shaping routine housework organization. Childcare policies are therefore a means of influencing what happens within the private sphere both directly—and policies aimed at fathers may reinforce their impact—and indirectly, by their effect on mothers' participation in employment and access to full-time employment.

However, childcare policies are only but one factor contributing to gender equality in the home. Analyzing the case of Portugal—in a rare comparison of southern European countries—Tavora (2012) showed that a high employment rate such as that in Portugal (one of the highest among European countries) can remain associated with very unequal sharing of housework(see Section 12.6.2). In addition to childcare policies, taxation and tax schemes—through their effect either on access to an independent income (de Henau et al., 2013) or on the labor market participation of a "second earner"—also influence how housework is shared within the household by changing the partners' relative resources. However, even though the sharing of housework depends on gender (in)equality in general (Fuwa, 2004), even in the most egalitarian countries, housework does not seem to be equally shared (Aassve et al., 2014; Geist and Cohen, 2011). Whether this is a case of lagged adaptation (Gershuny et al., 2005), incomplete revolution (Esping-Andersen, 2009), or stalled revolution (England, 2010) remains an open question. As for the impact of State policies, Dex (2010) underlines that the main changes seem to have produced policy changes rather than the contrary.

This section has examined gender inequality in nonmarket work, the other side of the considerable changes related to the rise in women's participation in the labor market and increased commitment to paid work over recent decades. On the one hand, these changes have generated tensions referred to as "second shift," "dual burden," and then, increasingly, "work–family balance" (or work–family reconciliation). On the other hand, they also result in increased economic independence, which has transformed women's opportunity sets within and outside the household and the economic foundations of marriage because there are fewer economic incentives to marry (see Lundberg and Pollak, 2013). These changes have also—to a greater or lesser extent in different countries—transformed the terms of gender (in)equality in general and the basis of intrahousehold allocations of time and money, with several broader social and economic consequences. Breen and Cooke (2005) proposed a game theoretical model in which threat points can be divorce or fertility within marriage; the outcome depends on whether men

[138] Another related issue, not addressed in this chapter, is that of eldercare, which has been variously found to affect women's participation in employment and working time (e.g., Kotsadam, 2011; Lilly et al., 2007; Spiess and Schneider, 2003). We have not examined it here because eldercare is rarely identified in time-use research (and often not measured as housework because it is often provided to people who do not live in the household).

prefer to live with an economically independent woman at the price of doing more housework or divorcing. Several researchers have pointed out the consequences of gender inequality in terms of fertility (e.g., Craig and Siminski, 2011; de Laat and Sevilla-Sanz, 2011; Haan and Wrohlich, 2011).

Picturing couples as battlefields, with each partner fighting to maintain or increase their relative power/earnings (possibly resulting in both of them overworking) is no more realistic than picturing them "doing gender" over the dishes. One possibility is that women have changed faster than contexts, norms, and conceptual tools. As underlined by Gregory (2009): "With women's progress in the labor market already undermining the basis of the Beckerian approach to the division of labor within the household, economic analysis will increasingly need a new paradigm for analyzing gender inequality" (p. 307). The same applies to sociological analysis.

It is a challenge, involving analysis of the difficult level of observation of the "intrahousehold" and conceptualization of the interactions between men and women within the household, between social norms and individual behaviors and contexts. The increasing availability of time-use data will be crucial for such progress, under two conditions: one, which is essential, is that time-use surveys include more than one respondent per household; the other is to find ways to improve non-time-use information, especially about incomes and wages, to escape the trade-off between reliable time-use information or reliable information on market wages.

12.7. WEALTH AND GENDER

Wealth inequality is an important dimension of economic inequality and an increasingly important driver of inequality (Piketty, 2013). Yet economic research has seldom considered wealth inequality between men and women, at least until recently[139]—so seldom that, in their review of empirical work on the distribution of wealth up to the late 1990s, Davies and Shorrocks (2000) cited only one study mentioning a gender differential (an analysis of the composition of wealth portfolios by Shorrocks, 1982). This reflects that research on wealth inequality considers essentially household wealth and hardly ever ownership within households. But does ownership matter within a household? Some assets such as a home or a car are a direct source of utility, whereas others, such as savings invested in interest-bearing assets, are a source of income; all are also a security—selling or drawing on one's savings helps to cope with variations in income. Then, one might answer that who owns what does not matter much because any member of the household can benefit from the assets held by the others—at least insofar as the household stays together—and the issue of ownership matters less than that of income sharing. But the distribution of assets between spouses may affect their relative power in decision making and the intrahousehold distribution of income, and household disruptions can make who owns what matter a great deal.

[139] See an overview of existing research, including in law and history, in Deere and Doss (2006).

What can be expected about a gender gap in wealth? In the perspective of wealth accumulation,[140] inequalities can result from inheritance and persistent differences in the levels and combined effects of incomes (which determine the capacity to save), preferences (which affect the savings rate), and investment choices (which determine the returns to savings). Considering the gender gap in earnings, the unequal labor market attachment of men and women, and a possible difference in investment behavior leading women to make less risky investments than men, resulting in lower returns to their financial savings, it is reasonable to expect a gender wealth gap in favor of men on average. However, because large proportions of men and women live in couples, there is also a "marital" dimension in the process of accumulation that somewhat complicates both the prediction and the assessment of the gender gap in wealth. As for transfers received, while inherited wealth is an important factor of wealth inequality in general, there is nothing in the empirical results examined to suggest a gender differential in the effect of inheritance (including expected inheritance).[141] But this is not the case for other types of transfers related to family disruptions such as divorce and widowhood, the latter "favoring" married women because they tend to outlive their husbands. This illustrates some of the complicated marital dimension of wealth accumulation and gender.[142]

This last section examines the main results from empirical work on gender inequality in wealth. Most of this empirical work is hampered by the lack of information about "who owns what" within households—with a few exceptions. As we will see first, data on individual wealth are imperfect and imprecise, a default that seriously limits the ability to assess the gender wealth gap. The section goes on to examine the main results of analyses of wealth inequality between men and women, and the last part reviews a few results on the distribution of wealth within couples.

[140] On models of accumulation see Davies and Shorrocks (2000).

[141] One could expect that it should not favor men (sons) any more or less than women (daughters), at least nowadays in most rich/Western countries; in continental Europe the prevalent regime is that of egalitarian inheritance (cf. Pestieau, 2003); in the United States and the United Kingdom there is freedom of bequest, but Cox (2003, p. 170) argues that the majority of bequests are shared equally between children. However, there could be less visible differences in gifts, financial support, and so on favoring sons more than daughters (see, e.g., Cox, 2003), but such inter vivos transfers are particularly not well covered by statistics.

[142] The effects of the marital dimension are complex. At the household level, living in a couple enhances the capacity to save and invest (especially in dual-earner couples) because of the economies of scale; therefore, a couple-household is likely to accumulate more wealth than two single households. At the individual level, between spouses/partners, wealth ownership is conditional to the marital status and premarital arrangement, which determine the degree of community/separateness of the assets, women's property rights, and divorce and inheritance law. In addition, "who owns what" may be influenced by the tax regime; dividing wealth within families to escape or reduce taxation may have contributed to the upward trend observed until the 1960s in women's share of wealth in the United States (Harbury and Hitchens, 1977), and, in countries where taxation is separate (e.g., Canada, the United Kingdom) it can be advantageous to divide the assets among spouses.

12.7.1 Who Owns What Within Households?

"Who" owns what within households is rather difficult to identify because, with very few exceptions, surveys collect information on household wealth[143]; surveys typically ask the respondent whether he/she/another member of the household owns such or such an item (and associated debt), then the value of each item if it was sold at the date of interview or elements allowing a value to be imputed.[144] Few surveys provide detailed information about the assets of all household members or whether the assets are owned individually or jointly.

The lack of data on individual wealth is probably one of the most serious limitations to research on wealth inequality between men and women. This is why most of the existing empirical work is based on household-level information and either compares men and women in different types of households (distinguishing single households headed by a man or a woman), divides the household wealth between its members, or combines partial information about ownership (when available) and computes a measure of "individualized" wealth. In our review of recent empirical work (1999–2013) addressing the question of the gender wealth gap,[145] we found very few studies able to make relatively precise individual imputations without making the assumption of equal ownership of all the assets.

In addition to surveys, there are other sources of information on wealth, either register data from taxes on wealth (when such taxes exist) or from estate tax records, but they are not really suitable for the analysis of gender inequality for several reasons. First, because of exemption thresholds, they only cover the upper part of the wealth distribution, possibly biasing estimates by gender if men and women are not evenly represented

[143] The most general approach to wealth is the "net worth", that is, the gross value of all the assets held minus associated debts. The assets generally covered include current bank accounts, saving accounts and saving plans, any type of financial product, life insurance, pension plans, owner-occupied residences and other real estate properties, business assets, durables (e.g., a car) and other tangible assets (such as art collections, jewelry, etc.). Debts include mortgage and other property debts, leasing contracts, consumer credit, etc. In most studies, these items are aggregated in four groups: financial assets, housing wealth, business wealth, and pension wealth.

[144] Measurement error caused by nonresponse and misreporting is an important issue (see detailed reviews by Davies and Shorrocks, 2000; Juster and Kuester, 1991); respondents may be reluctant to disclose information, omit reporting some assets or understate their value, or have difficulties in assessing the value of the assets held (e.g., a house, a financial portfolio). To help with this assessment, questionnaires often allow the respondent to report bracketed values instead of amounts—hence a choice between lack of precision and missing responses (see Chand and Gan, 2003). To avoid "don't know" answers from respondents who have little or no knowledge of the assets held in the household (or by whom, if this information is collected), surveys often start by identifying the household member who is most likely to have such knowledge (the "best financial respondent" or "best knowledgeable respondent").

[145] Thus excluding studies that consider only women's wealth (e.g., Gornick et al., 2009; Sanders and Porterfield, 2010). Also striking in this review is the small number of studies on non-US empirical work.

over the distribution (survey data, on the contrary, tend to underrepresent this upper part, but this is generally corrected by oversampling the top of the distribution). In addition, taxes on wealth are levied on fiscal units that are not necessarily individuals. The major drawback of estate tax records, which provide information on individual wealth holdings, is that the information only covers dead people. However, using estimates derived from this information,[146] a small number of studies from the 1970s and early 1980s provide some evidence on the distribution of wealth by gender in the United Kingdom and the United States.[147] They observed that women held less wealth than men but that there was a larger gap at the top of the wealth distribution (Atkinson, 1971; Atkinson and Harrison, 1978; Smith, 1974), a greater concentration of wealth among women than among men (Atkinson and Harrison, 1978), differences in the composition of assets, fewer debts in the richest women's wealth than in the richest men's (Smith, 1974), differences in the origin of the assets (Harbury and Hitchens, 1977), and differences in the influence of age and wealth on the composition of men's and women's portfolios (Shorrocks, 1982). However, these results, while they provide valuable indications of differing patterns of wealth accumulation (albeit among a population limited to the wealthiest), describe generations from before "the quiet revolution" (the estimates are derived from estate data in the 1960s and 1970s), when women's wealth was more likely to depend on their fathers' or husbands' wealth than on "self-made" wealth (see Harbury and Hitchens, 1977), fewer women had independent incomes, and marriages were more stable than today. In the case of the United States, a recent estate-based study by Edlund and Kopczuk (2009) shed some doubt about the actual effect of the change in women's economic status. They showed that the share of women among the very wealthy, after peaking at the end of the 1960s, returned to its prewar level by 2000. Their analysis of this evolution is that the share of women at the top of the wealth distribution reflects the share of inherited wealth in an inverse relation with technological change: the information technology revolution (in which men made new fortunes) coincides with the turning point, in the 1970s, in the share of women among the very wealthy. Among the most wealthy, the trends Edlund and Kopczuk outlined ultimately point out the continuing gender gap in entrepreneurship.

[146] These methods (mortality multipliers and estate multipliers) basically expand the information on the deceased to estimate wealth holdings among the living. These methods are presented and discussed by Atkinson and Harrison (1978) and Davies and Shorrocks (2000).

[147] These studies aimed firstly to analyze the distribution of wealth, its concentration and intergenerational transfers. In a way, the data obliged researchers to consider men and women separately. One of the problems was to obtain estimates of family wealth starting from information on men's and women's individual wealth (see Atkinson and Harrison, 1978, p. 241), quite the opposite of the problem research on gender inequality faces today, which is to separate men's and women's wealth starting from information on household wealth.

12.7.2 Investigating the Gender Gap in Wealth

Facing the unanswered question of "whose asset is it?" in couple households (and multi-person households in general), researchers have adopted strategies that, one way or another, sidestep the issue of the distribution of wealth within the household. We first describe these strategies and their limitations, then the main results.

12.7.2.1 Strategies and Limitations

There are two main ways to avoid the issue of intrahousehold distribution of assets. One (the most frequent) is to compare the net worth of men or women living in couple households and those in "not-couple" households (i.e., households in which the head/reference person has no partner or spouse at the time of survey) and use the gender and other characteristics of the household's head as explanatory variables of the household's wealth. Another strategy consists of imputing "individualized" levels of wealth (either per capita net worth, assigning equal shares to each partner in couple households, or using combined estimates, adding individual assets for which individual-level information exists and equal shares of the other assets[148]). Alternatively, some studies focus on one sole type of asset for which their data provide individual-level information. The selection of studies referred to in the review of results is presented in Table 12.6.

While better than nothing, none of these options is satisfying. First, comparing couples and single individuals obviously conflates gender and family/household composition. Most studies focus on the gap between single men and women heading households, but men and women are likely to head different sorts of households[149]—most households headed by women contain no adult man, and single men heading households are more likely to be young and/or childless. Households headed by single women can also be more heterogeneous than those headed by single men, as underlined by Conley and Ryvicker (2005). Single women heading households are more likely to have children or to be older (and widowed), hence conflating not only gender and family composition but also life cycle effects. The "individualization" option is very likely to result in biased estimates of individuals' wealth holdings.[150] For couple households,

[148] This methodology is close to that used by Meulders and O'Dorchai (2010) to compute individual incomes (see Section 12.2).

[149] Despite cross-country differences, the determination of the head of household in surveys is such that it results most often in identifying a man if there is one in the household (see the review of statistical practices by Cowell et al., 2012).

[150] Per capita estimates raise the issue of a pertinent equivalence scale to use in comparisons of wealth among households of different size; see the discussion by Sierminska and Smeeding (2005). Because studies based on household-level information compare households of different sizes, using the household net worth or a per capita or other individualized measure should reduce the wealth gap between couples and other types of households, depending on the size of the other households. Schmidt and Sevak (2006) used a variant of the per capita measure that actually affects the level and significance of the coefficients; they mentioned having tested various alternative equivalence scales with similar results.

Table 12.6 Research strategy, units of analysis, and wealth measure

Author (year)	Country, data source, years	Units of analysis (age group)	Wealth components	Wealth measure
A. Household wealth				
Conley and Ryvicker (2005)	United States, PSID, 1984–1989	Single heads of household (aged ≥18)	Pension wealth excluded	Household net worth
Lupton and Smith (2003)	United States, HRS wave 1992 (cohort born 1931–1941)	Couples versus single (aged ≥18)	Total net worth (not pension wealth)	Household net worth
	United States, PSID, 1984–1989–1994			
Ozawa and Lee (2006)	United States, SCF, 1998	Couples versus single heads of household (age 25–61)	Household net worth	
Sedo and Kossoudji (2004)	United States, SIPP, 1996	Individuals (reference person for the housing) (age ≥25	Only housing wealth Household housing wealth	Gross value and net worth
Schmidt and Sevak (2006)	United States, PSID, 2001	Individuals (married versus single) (age ≥18 and the younger cohort age 25–39)	Pension wealth excluded	Household net worth Variant: per capita
Ruel and Hauser (2013)	United States, WSL wave 2004	Individuals (best financial reporter), cohort graduating in 1957		Household net worth
Ulker (2009)	United States, HRS wave 1992	Couples versus single heads of household (cohort born 1931–1941)		Household net worth
Yamokoski and Keister (2006) US, *NSLY1979*		Couples versus single heads of household (cohort born 1957–1964)		Household net worth
B. Household wealth individualized				
Denton and Boos (2007)	Canada, SFS, 1999	Individuals (age ≥45)		Per capita net worth
Warren (2006)	United Kingdom, FRS, 1995–1996	Individuals (age ≥18)		Single: Household net worth Couple: own pension wealth + half the couple's financial and housing net worth
Wilmoth and Koso (2002)	United States, HRS wave 1992	Individuals (cohort born 1931–1941)		Per capita net worth

FRS, Family Resources Survey; HRS, Health and Retirement Survey; NLSY1979, National Longitudinal Survey of Youth 1979 cohort; PSID, Panel Survey of Income Dynamics; SCF, Survey of Consumer Finances; SFS, Survey of Financial Security; SIPP, Survey of Income and Program Participation; WLS, Wisconsin Longitudinal Survey.

dividing the couple's assets equally entails the assumption that all assets are owned jointly and that each partner holds an equal share. This is a strong assumption in the case of cohabiting couples and a very simplifying assumption in the case of married couples; some assets may indeed be equally owned, but there is no reason for joint ownership to be systematic, and the partners' shares of assets may differ depending on the marital status and marriage agreement. Hence, allocating couples' wealth between spouses under the assumption of joint ownership and equal share of all assets is very likely to conceal within-household inequality and, in turn, to bias the comparison between men and women living in different types of households. Frick et al. (2007), using appropriate data, showed that wealth is more unequally distributed between individuals when measured on the basis of individual-level information than when measured per capita on the basis of household-level information.

Finally, comparing men's and women's wealth over only one type of asset obviously provides a partial measure of the extent of gender wealth inequality. It is an attractive option for analytical purposes (e.g., selection in ownership; cf. Sedo and Kossoudji, 2004), but it does not account for the choice between alternative investments or the influence of total wealth and family composition on the composition of assets.

This review of obstacles and limitations may seem somewhat discouraging. However, working within these limits provides some indications of the influence of marital status and history in the process of wealth accumulation between men and women and points to some gender differences in wealth ownership. For simplicity, we talk of "married" men and women for couple households (including cohabiting couples unless results distinguish between marital status) and of "single" men and women for single households.

12.7.2.2 Evidence: Gender and Composition Effects

As might be expected, the result regularly obtained is that, once relevant characteristics have been controlled for (the minimum set includes age, education, inheritance, race, a measure of income, and/or current labor market status[151]), married men and women have an advantage in the accumulation of wealth over single men and women, that is, a "marriage effect" (Denton and Boos, 2007; Lupton and Smith, 2003; Ozawa and Lee,

[151] Assessing the impact of gender/household composition on the wealth differential leads to problems of estimation. In particular, income, labor market status, marital status, and marital history can be influenced by the ownership of assets and the level of wealth. For income, a number of studies use only the earnings, nonasset income, or a proxy for "permanent income." Issues of endogeneity are mentioned in relation to divorce (Ulker, 2009) or marriage (Ruel and Hauser, 2013) but are not dealt with because of data limitations. Wealth is also much more unequally distributed than income or earnings (as illustrated by the large gaps between the mean and median values of wealth holdings when descriptive statistics on the dependent variable are provided), resulting in other biases that are not systematically acknowledged. Many empirical results face one of these econometric problems. We have chosen not to dwell on these limitations.

2006; Schmidt and Sevak, 2006; Wilmoth and Koso, 2002; Yamokoski and Keister, 2006). The second general result is that single men have an advantage over single women (Conley and Ryvicker, 2005; Denton and Boos, 2007; Lupton and Smith, 2003; Ozawa and Lee, 2006; Schmidt and Sevak, 2006; Wilmoth and Koso, 2002). Finally, virtually all the results point to single parents (mostly women) as the most disadvantaged group compared with any other combination of gender and household composition.

Beyond this general hierarchy related to household composition, there are some discrepancies. One is between wealth levels: the marriage effect is larger in the upper part of the distribution, and the gap between single men and single women disappears at the 75th percentile (Schmidt and Sevak, 2006). Using a measure of savings, Lupton and Smith (2003) found a larger advantage of married couples at higher levels of net worth and discuss a possible impact of sorting into marriage (i.e., "prudent" individuals would be more likely to marry and to stay married or to remarry after a divorce because they also find that divorced and remarried individuals have better outcomes than divorced people).

The other discrepancy is between cohorts. When they restrict their sample to a cohort aged between 25 and 39 in the early 2000s, Schmidt and Sevak (2006) no longer obtained statistically different effects of household composition or gender, indicating either a cohort effect or a life cycle effect. In the first case, this could result from the closing of other gender gaps; in the second, it would indicate that, because wealth inequality builds over time, these young households have not yet accumulated enough for a gender and a family differential to appear. These results do not entirely converge with those obtained from another cohort of about the same age at the same period: Yamokoski and Keister (2006) found almost no difference between never-married men and women and no significant difference between divorced women and men with children, but they found a strong marriage effect. There is also a marriage effect among older cohorts, although it is mitigated by marital history. In short, any disruption from continuous marriage has a negative impact on accumulated wealth and any past marital dissolution has a negative impact (Ulker, 2009; Wilmoth and Koso, 2002; both used data on the wealth of a cohort born between 1931 and 1941 and collected in 1992). However, the effects differ for men and women depending on the type of disruption and the current marital status; compared with continuously married men, the effect of having never married, having divorced once or twice, or being separated after a second marriage (the worst case) are significantly larger for women than for men (Wilmoth and Koso, 2002).

Ruel and Hauser (2013) studied a cohort of men and women who graduated in 1957 (wealth was measured in 2004). Among those not currently married, they found no real gender effect on the accumulated net worth—the difference between men and women is essentially explained by labor market history, past income (nonasset income), and current earnings (controlling for education, social origin, and inheritances). Nevertheless, divorce has a small negative effect on women's net worth, but the major factor is the gender earnings gap over the work-life cycle. Among the currently married, and compared with continuously married men, continuously married women have a significant disadvantage, as do divorced men and, even more so, divorced women. But their full model explains a rather small portion of the accumulated wealth, suggesting unobserved heterogeneity (e.g., different preferences in savings and investment). For currently married men and women, they discussed a possible effect of comparing men and women as the "best financial respondent": if men are more likely to be the couple's best respondent in wealthy couples and women in less wealthy couples, the gender of the respondent depends on the household's level of wealth.[152]

In addition to the effect of the interaction of gender and household composition, other studies have investigated more specifically the effect of the labor market dimension. Women's lower earnings affect their capacity to save, and they have less continuous careers. They are therefore likely to accumulate fewer assets than men (Denton and Boos, 2007) and less pension savings (Warren, 2006). Other occupational differences also determine different opportunities to accumulate assets. For example men, who are more often self-employed than women (see Section 12.4), are more likely to hold business assets. In the United States in 1998, households headed by single men owned about three times the net worth owned by those headed by single women (Ozawa and Lee, 2006). In Germany in 2002, the ratio of men's to women's business assets (gross value) was about 5.5 (Sierminska et al., 2010). In their full model of married men and women, Ruel and Hauser (2013) obtained large and significant coefficients for self-employed men (among the largest coefficients, together with past incomes).

Empirical work based on household-level information on wealth suffers from several limitations but provides two main results: virtually all the results show strong effects of marital status (understood as a type of household/family)—but mitigated effects of marital history—and significant differences related to labor market outcomes, that is, the two essential dimensions of gender economic inequality, but not independent of each other. The main limitation is that the results are unclear about the exact unit analyzed because the dependent variable is a household variable and the independent variables are those of one person in the household.[153] This mixing of gender and marital status/household composition and finally parenthood (a second best justified by the absence of appropriate data) has many limits.

[152] This would be consistent with findings on money management within households, showing that women are more likely to manage tight resources (e.g., Kenney, 2006; Vogler et al., 2008). This raises an issue of measurement error if, for whatever reason, there is a systematic gender bias, as suggested by Zagorsky (2003). Using US data, Zagorsky found significant differences in the value of a couple's net worth reported by the wife or her husband because of husbands reporting a higher gross value of the assets than their wives and wives reporting higher levels of debt than their husbands.

[153] This problem is very similar to that underlined in Section 12.2 about the working poor.

12.7.3 Measuring the Gender Gap in Wealth and Its Components

So far, only two studies, those by Sierminska et al. (2010) and Bonnet et al. (2013), have measured and investigated the gender wealth gap using data that allow a gender gap in wealth to be computed, that is, considering men and women in any type of household. While the data they use (the German Socio-Economic Panel 2002 and the French Household Wealth Survey "Enquête Patrimoines" 2003–2004, respectively) do not provide directly individual wealth variables, they do provide information allowing individual wealth to be computed (on the basis of questions about the share of common property held by couples' partners) rather than assuming equal ownership. Both these studies found a substantial gender wealth gap: men's net worth is higher than women's by 45% in Germany and by 16% in France. Accounting for at least some of this huge difference between the two countries, Bonnet et al. did not include business assets in their definition of wealth; in Germany, the gender gap in business wealth is the highest of the gaps by component[154] (the men-to-women ratio is 5.5). In particular, the total gender gap is higher in the subsample of married couples than on average in Germany (similar in France), higher in the subsample of cohabitant partners than among the married subsample—suggesting different financial arrangements—and lower than the average in the subsample of singles living alone (and nonexistent in France in 2009–2010).

OLS regressions show all the expected associations between wealth, marital status, and labor market characteristics, controlling for education, permanent income, and social origin.[155] The most interesting results are, for both countries, the decomposition of the gender wealth gap, even more so because the two studies use the same methodology (DiNardo et al., 1996), although they are not fully comparable because Sierminska et al. (2010) only analyzed the composition of the gap in their sample restricted to married/cohabiting men and women, and Bonnet et al. (2013) decomposed the gap of their full sample of men and women. In both countries, the main contributions to the gender wealth gap were income and a set of labor market characteristics, especially at the median and upper deciles of the distribution.[156] But the most striking result is the relatively large share of the gender wealth gap that is not explained by the characteristics, especially in the lower part of the distribution (and larger in Germany than in France) and its sign (negative) in both countries.[157] These results suggest that, with given characteristics, women obtain

[154] The self-employed not only hold business assets that employees do not accumulate but also invest more in private pensions because they are not covered by the social security regime.

[155] Using a pooled sample of men and women, Bonnet et al. (2013) obtained very comparable results in 2003–2004 and 2009–2010; the sign and significance level of almost all variables remains the same, except for "cohabiting," which was significant only in 2009–2010.

[156] The 75th percentile in the decomposition for Germany (Sierminska et al., 2010) is puzzling; contrary to any other point, education and intergenerational characteristics are negative and the unexplained gap is positive.

[157] This result is obtained at any point of the distribution in France but not in the lower half of the upper distribution in Germany (the restriction of the sample could account for this difference).

more wealth than men, that is, they get better returns to their characteristics. But men have, on average, "better" characteristics than women and therefore own more wealth.

In both papers, the authors speculate about these "better returns" along two main tracks. One is the "marital" track: women benefit from the relatively better characteristics of their partner/husband (and perhaps deceased husband); this line, consistent with the "male breadwinner model," is especially attractive for Germany (the results are for married men and women); the results for France do not allow anything more than speculation, but it could be less relevant given that one of the major differences between the two countries is the higher degree of gender inequality in terms of labor market characteristics and outcomes in Germany.

The other line of inquiry concerns gender differences in "preferences," involving risk preferences and their effect on savings and investment choices. Risk preference determines the share allocated to risky assets, which are assumed to yield better returns; income depends on the allocation of savings to more or less risky investments and in turn determines the potential for future investment. A gender difference in risk preference could then, in the long run, help to explain a wealth gap between men and women who are otherwise comparable. The question of a gender effect on risk tolerance/aversion has been investigated extensively (see Section 12.3), providing indications that women are more risk averse than men or that men are overconfident (e.g., Barber and Odean, 2001). However, risk preferences might explain very little of the gender gap in wealth. Estimating the choice of risky assets (the share of stocks in older Americans' individual retirement accounts), Neelakantan (2010) found that the gender differential in risk tolerance accounts for 10% and the gender gap in earnings for 51% of the difference between men's and women's accumulated retirement accounts. Barasinska and Schäfer (2013) also showed that between men and women with similar risk preferences (measured by the presence of risky assets in their financial portfolio), there is no significant difference in the allocation decision (measured by the share of risky financial investments), except in Italy—a result that they related to a choice of identity in a context of high gender inequality. Other research invokes a possible effect of a differential in financial literacy (e.g., Van Rooij et al., 2011). Dwyer et al. (2002) found that controlling for financial knowledge significantly reduced the impact of gender on investment decisions but that this effect was difficult to assess among representative samples.

The effect of gender differences in the choice of investment could be more complicated to assess in the case of men and women living together (a large share of the adult population); savings and investment decisions may result from the interaction of partners' preferences. Finally, the issue of investment choice raises the issue of wealth inequality and power in decision making within the household.

12.7.4 Wealth and Gender Within the Household

Given the lack of information on individual wealth, intrahousehold (intracouple) wealth inequality and the role of partners' preferences in saving and investing decision making have seldom been analyzed. A first question is that of the pattern of ownership (joint/separate) and the distribution of wealth between spouses/partners. Grabka et al. (2013), using the German Socio-Economic Panel 2007, found a large within-couple gender wealth gap; on average, women's assets represent about 37% of the couple's wealth—a quite different picture from the assumed equality—and the gap tends to increase with the wealth level. Men hold more wealth than women in 52% of couples, women hold more in 29% of couples, and the shares are equal in the remaining 19%. Kan and Laurie (2013), using the British Household Panel Survey (HPS) (which collected data on savings, financial investments, and debts at the individual level in 1995, 2000, and 2005), found that savings are more often jointly held than investments or debts and, controlling for ownership, that the likelihood of joint ownership of the three types of assets is lower among cohabiting couples—a pattern already identified for income pooling (e.g., Vogler et al., 2006; see also Section 12.2).[158]

The next issue concerns the relation between gender, preferences, financial decision making, and wealth within couples. The general pattern that emerges from various empirical results is that there is a relation, but it is mediated by the level of income, wealth, and context. In the Netherlands, Barasinska and Schäfer (2013) found that in couples who reported that they make financial decisions jointly, gender has no effect on the probability of ownership of risky assets but the differential in risk tolerance between the spouses has an effect: any level of risk tolerance of one spouse has a reduced effect if the other spouse is more risk averse. Such an effect of mitigation is illustrated by Love (2010). Studying allocations after a divorce, he showed that men choose riskier investments than women and that transitions from divorce to marriage have opposite effects. In the United States, Neelakantan et al. (2009) found support for the prediction of the collective model: the household portfolio is determined by the risk preference of the spouse with more bargaining power. This result is somewhat contradicted by Gibson et al. (2006) in Australia; they found that the level of accumulated wealth among preretirement couples is not higher when women have more bargaining power (measured by a power index combining age, education, inherited wealth, and income); the reason is that the public pension system in Australia better replaces preretirement income for women than for men. In Canada, Phipps and Woolley (2008), using various measures of control over money, found that women's greater control is associated with lower probabilities of contributions to a private savings plan both for men and women. However, women's control is more frequent at lower-income levels, and savings are primarily associated with income. Finally, in Germany, Grabka et al. (2013) found that the wealth gap tends to be higher in couples where the financial decisions are most often made by the man and that men are more likely to make financial decisions in wealthier couples (a question also

[158] They also observed a downward trend in joint ownership between 1995 and 2005, consistent with the trend in marriage versus cohabitation observed in the United Kingdom and other countries.

addressed by Ruel and Hauser, 2013). All in all, things are rather difficult to disentangle—especially because joint ownership is almost never controlled for, "who makes decisions" is self-reported, and there may be divergence between each spouse's view of "who decides" (for retired couples in the United States see Elder and Rudolph, 2003) or about the couple's financial difficulties (see Breunig et al., 2007, with Australian data). Further research taking into account a possible difference between what happens in married *and* cohabiting couples would also be useful.

What conclusions can be drawn from research on the gender wealth gap in the past decade? First, it seems that there really is a gender gap in wealth, but there are far too few empirical studies that can provide an assessment of its full extent. For the time being, one could say that the gender gap in wealth seems easier to explain than to measure. As for the explanations, they converge on a strong effect of gender earnings differentials mitigated by strong effects of marital status and history, complicated by additional effects of cohorts and total wealth. In addition, a large share of the existing empirical work consists of research on the United States, with a particular household structure[159] and based on inappropriate data. It is an understatement to conclude that more research—and comparative work—is needed to allow for common methodologies to build up.

Identifying who owns and controls assets within households is a prerequisite to gaining knowledge that has important implications for the analysis of gender inequality, as underlined by Deere and Doss (2006). Gaining knowledge is conditioned by the availability of appropriate data. This is a challenge for statistics, a challenge seldom acknowledged to date, as exemplified by the recent initiative of the European Central Bank aimed at developing harmonized data on wealth. On the model of the Canberra Group, the Central Bank seems to have adopted the standard unitary approach; the questionnaires available online[160] show no intention to collect any details on individuals or ownership status. The Luxembourg Wealth Study,[161] a project aimed at building a data set based on harmonized national data, seems to have missed the issue, too (see Barasinska and Schäfer, 2013, who mention that individual information is lost through data standardization [p. 8]). As mentioned in Section 12.2 in our discussion of income statistics, the increased financial independence of women and changing patterns of families and households provide good reasons to argue that statistics must change as well.

12.8. CONCLUSION

Gender economic inequality encompasses a large field of study, far beyond the gender wage gap (even though this issue is central); the related literature is overabundant and

[159] As shown by Bover (2010) in a comparison between Spain and the United States.
[160] http://www.ecb.europa.eu/home/html/researcher_hfcn.en.html.
[161] Luxembourg Wealth Study (2003); http://www.lisdatacenter.org/our-data/lws-database/.

still growing, driven by the conjunction of dramatic demographic and economic shifts in the past decades and the puzzle of persistent gender inequalities. Delayed maternity and marriage have transformed women's economic opportunities relative to men's: education is no longer a key factor of gender inequality, and motherhood is less incompatible with paid work. Family structures have changed in several ways. A majority of couple families count two earners, but marriages have become less stable than in the past, and cohabitation is an increasing arrangement. Consequently, women may rely less on a husband's incomes, as in the traditional "male breadwinner" model, and more on own their economic independence.

But the "quiet revolution" does not seem to have changed market and nonmarket work to a comparable extent, and the gender wage gap has remained basically unchanged in the past two decades. More than an academic puzzle, this is also a major policy issue. There is a general consensus that this persistence is related to occupational segregation and a gender gap in promotions, but the underlying causes are still debated. Whether the explanation focuses on employer discrimination, gender psychological differences and social norms that disadvantage women in the labor market, or family constraints on work hours, a common denominator is the gender division of labor, especially of unpaid childcare, as reflected by the importance given to the issue of work–family balance. As Craig (2006) remarks: "In the absence of adequate support, there is a sticking point in the revolution: taking care of the kids. An implication of this is that the marker of the most extreme difference in life opportunities between men and women may not be gender itself, but gender combined with parenthood" (p. 146).

While the time dedicated to housework has decreased dramatically since the 1960s, women still make the major contribution to nonmarket work within households, and the need for flexible work hours when they are mothers affect their labor market outcomes in many ways, from less favorable careers to smaller pensions—a delayed expression of the gender asymmetrical effects of family life. So, it is not surprising that recent public policies addressing gender wage inequality point out the need for making men and women more equal in terms of family constraints, adding incentives to get fathers more involved in childcare to usual measures of work–family reconciliation for mothers.

All throughout this chapter, various limitations in the knowledge of gender inequality have been underlined, pointing at the lack of appropriate data on individual incomes. Work incomes are quite well identified at individual level, but this is not so for other income components, which in many data sets are available only at the household level. It is quite the opposite with time-use surveys, which provide individual-level information but do not always allow reliable household-level variables to be computed because very often there is only one respondent per household. Another problem with data is that the possibilities for researching the relations between time use, labor market outcomes, and issues of power within the household are limited. Time-use data, on the one hand, are a precious resource for analyzing the allocation of time, but variables describing earnings and other incomes often lack of precision, whereas data on income and labor market status, on the other hand, most often provide imprecise—if any—information on the time not spent at work. Finally on wealth, an emerging strand of research on gender inequality, the need for appropriate statistical sources is blatant. Cross-national perspectives are needed also to better understand the various dimensions of gender economic inequality. This entails discussing and adopting common methodologies allowing meaningful international comparisons - and harmonization is clearly an issue, but also the agreement on basic principles, especially of the pertinent level of information.

A better understanding of gender economic inequality is not only a question of data (although more large scale and comparable data would not hurt), but also a question of concepts and indicators allowing to compare men and women economic outcomes. Women's economic and social status have undergone enormous changes over the last decades, and this is challenging many theoretical approaches. A concern is that the reference remains of the household as a stable unit of pooling and sharing, an increasingly irrelevant conceptualization while households are less stable than in the past and individuals are more likely to experience diverse family configurations over a lifecycle. These are serious incentives to depart from a paradigm which limits the analysis of inequality between individuals in general and between men and women in particular.

REFERENCES

Aassve, A., Fuochi, G., Mencarini, L., 2014. Desperate housework relative resources, time availability, economic dependency, and gender ideology across Europe. J. Fam. Issues 35 (8), 1000–1022.

Adams, R., Gray, S., Nowland, J., 2011. Does gender matter in the boardroom? Evidence from the market reaction to mandatory new director announcements. Working Paper.

Adelman, L., Middleton, S., Ashworth, K., 2002. Intra-household distribution of poverty and social exclusion: evidence from the 1999 PSE survey of Britain. Centre for Research in Social Policy, Working Paper No. 23.

Agarwal, B., 1997. Bargaining and gender relations: within and beyond the household. Fem. Econ. 3 (1), 1–51.

Aggarwal, R., Goodell, J.W., 2013. Political-economy of pension plans: impact of institutions, gender, and culture. J. Bank. Financ. 37 (6), 1860–1879.

Aguiar, M., Hurst, E., 2008. The increase in leisure inequality. National Bureau of Economic Research Working Paper No. w13837.

Ahern, K.R., Dittmar, A.K., 2012. The changing of the boards: the impact on firm valuation of mandated female board representation. Q. J. Econ. 127 (1), 137–197.

Ahmad, N., Koh, S., 2011. Incorporating estimates of household production of non-market services into international comparisons of material well-being. OECD Statistics Working Papers, 2011/07 OECD Publishing, http://dx.doi.org/10.1787/5kg3h0jgk87g-en.

Ahmad, N., Seymour, R., 2008. Defining entrepreneurial activity: definitions supporting frameworks for data collection. Working Paper, STD/DOC(2008)/2, OECD.

Aigner, D.J., Cain, G.G., 1977. Statistical theories of discrimination in labor markets. Ind. Labor Relat. Rev. 30 (2), 175–187.

Akabayashi, H., 2006. The labor supply of married women and spousal tax deductions in Japan—a structural estimation. Rev. Econ. Househ. 4 (4), 349–378.

Akerlof, G.A., Kranton, R.E., 2000. Economics and identity. Q. J. Econ. 115 (3), 715–753.

Akerlof, G.A., Kranton, R.E., 2010. Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being. Princeton University Press, Princeton.

Albanesi, S., Olivetti, C., 2009. Home production, market production and the gender wage gap: incentives and expectations. Rev. Econ. Dyn. 12 (1), 80–107.

Albanesi, S., Sahin, A., 2013. The gender unemployment gap. Research Paper, No. 613, Federal Reserve Bank of New York.

Albrecht, J.W., Edin, P.-E., Sundström, M., Vroman, S.B., 1999. Career interruptions and subsequent earnings: a reexamination using Swedish data. J. Hum. Resour. 34 (2), 294–311.

Albrecht, J., Björklund, A., Vroman, S., 2003. Is there a glass ceiling in Sweden? J. Labour Econ. 21 (1), 145–177.

Alderman, H., Chiappori, P.-A., Haddad, L., Hoddinott, J., Kanbur, R., 1995. Unitary versus collective models of the household: is it time to shift the burden of proof? World Bank Res. Obs. 10 (1), 1–19.

Alesina, A., Ichino, A., Karabarbounis, L., 2011. Gender-based taxation and the division of family chores. Am. Econ. J. 3 (2), 1–40.

Alesina, A.F., Lotti, F., Mistrulli, P.E., 2013. Do women pay more for credit? Evidence from Italy. J. Eur. Econ. Assoc. 11 (s1), 45–66.

Allard, M.D., Bianchi, S., Stewart, J., Wight, V.R., 2007. Comparing childcare measures in the ATUS and earlier time-diary studies. Mon. Labor Rev. 130, 27–36.

Allen, I.E., Langowitz, N.S., 2013. Understanding the gender gap in entrepreneurship: a multicountry examination. In: The Dynamics of Entrepreneurship: Evidence from Global Entrepreneurship Monitor Data. Oxford University Press, Oxford, pp. 31–57.

Altonji, J.G., Blank, R.M., 1999. Race and gender in the labor market. In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics. vol. 3. Elsevier Science B.V., Amsterdam, pp. 3143–3259.

Alvarez, B., Miles, D., 2003. Gender effect on housework allocation: evidence from Spanish two-earner couples. J. Popul. Econ. 16 (2), 227–242.

Amilon, A., 2012. The effects of a Danish legal reform on divorce probabilities and pension savings. Rev. Econ. Househ. Jan, 1–26.

Amuedo-Dorantes, C., De la Rica, S., 2006. The role of segregation and pay structure on the gender wage gap: evidence from matched employer–employee data for Spain. Contrib. Econ. Anal. Policy 5 (1), 1498–1523.

Anderson, P., Wadensjo, E., 2008. A gender perspective on self-employment entry and performance as self-employed. IZA Discussion Paper no. 3581.

Anderson, D.J., Binder, M., Krause, K., 2002. The motherhood wage penalty: which mothers pay it and why? Am. Econ. Rev. 92 (2), 354–358.

Anderson, D.J., Binder, M., Krause, K., 2003. The motherhood wage penalty revisited: experience, heterogeneity, work effort, and work-schedule flexibility. Ind. Labor Relat. Rev. 56, 273–294.

Andrieux, V., Chantel, C., 2011. Les retraites perçues fin 2008. Etudes et résultats N°758, Drees.

Anker, R., 1998. Gender and Jobs: Sex Segregation of Occupations in the World. International Labour Office, Geneva.

Antecol, H., 2011. The opt-out revolution: Recent trends in female labor supply. Research in labor economics 33, 45–83.

Antecol, H., Cobb-Clark, D.A., 2013. Do psychosocial traits help explain gender segregation in young people's occupations? Labour Econ. 21, 59–73.

Antonovics, K., Town, R., 2004. Are all the good men married? Uncovering the sources of the marital wage premium. Am. Econ. Rev. 94 (2), 317–321.

Anxo, D., Mencarini, L., Pailhé, A., Solaz, A., Tanturri, M.L., Flood, L., 2011. Gender differences in time use over the life course in France, Italy, Sweden, and the US. Fem. Econ. 17 (3), 159–195.

Apps, P.F., 2003. Gender, time use and models of the household. IZA DP No. 796.

Apps, P.F., Rees, R., 1988. Taxation and the household. J. Public Econ. 35, 355–369.

Apps, P.F., Rees, R., 1997. Collective labor supply and household production. J. Polit. Econ. 105 (1), 178–190.

Apps, P.F., Rees, R., 1999. Individual versus joint taxation in models with household production. J. Polit. Econ. 107 (2), 393–403.

Apps, P.F., Rees, R., 2002. Household production, full consumption and the costs of children. Labour Econ. 8 (6), 621–648.

Apps, P.F., Rees, R., 2007. The taxation of couples. IZA Discussion Papers No. 2910.

Arrow, K., 1973. The theory of discrimination. Working Paper No. 30(A), Princeton University.

Arulampalam, W., Booth, A.L., Bryan, M.L., 2007. Is there a glass ceiling over Europe? Exploring the gender pay gap across the wage distribution. Ind. Labor Relat. Rev. 60 (2), 163–186.

Ashby, K.J., Burgoyne, C.B., 2008. Separate financial entities? Beyond categories of money management. J. Socio-Econ. 37 (2), 458–480.

Asiedu, E., Freeman, J.A., Nti-Addae, A., 2012. Access to credit by small businesses: how relevant are race, ethnicity, and gender? Am. Econ. Rev. 102 (3), 532–537.

Aslaksen, I., Koren, C., 1996. Unpaid household work and the distribution of extended income: the Norwegian experience. Fem. Econ. 2 (3), 65–80.

Atkinson, A.B., 1971. The distribution of wealth and the individual life cycle. Oxf. Econ. Pap. 23 (2), 239–254.

Atkinson, A.B., Harrison, A.J., 1978. Distribution of Personal Wealth in Britain. CUP Archive, Cambridge.

Attanasio, O., Lechene, V., 2002. Tests of income pooling in household decisions. Rev. Econ. Dyn. 5 (4), 720–748.

Attanasio, O., Low, H., Sanchez-Marcos, V., 2008. Explaining changes in female labor supply in a life-cycle model. Am. Econ. Rev. 98 (4), 1517–1552.

Azmat, G., Petrongolo, B., 2014. Gender and the labor market: what have we learned from field and lab experiments? WP40 CEPR.

Azmat, G., Güell, M., Manning, A., 2006. Gender gaps in unemployment rates in OECD countries. J. Labor Econ. 24 (1), 1–37.

Babcock, L., Laschever, S., 2003. Women Don't Ask. Negotiation and the Gender Divide. Princeton University Press, Princeton, NJ.

Bagues, M., Esteve-Volart, B., 2010. Can gender parity break the glass ceiling? Evidence from a repeated randomized experiment. Rev. Econ. Stud. 77, 1301–1328.

Bajtelsmit, V., 2006. Gender, the family, and economy. In: Clark, G.L., Munnell, A.H., Orszag, J.M. (Eds.), The Oxford Handbook of Pensions and Retirement Income. Oxford University Press, pp. 121–140.

Baker, M., Fortin, N.M., 1999. Women's wages in women's work: a US/Canada comparison of the roles of unions and "public goods" sector jobs. Am. Econ. Rev. 89 (2), 198–203.

Baker, M., Milligan, K., 2008. How does job-protected maternity leave affect mothers' employment? J. Labor Econ. 26 (4), 655–691.

Baker, M., Gruber, J., Milligan, K., 2005. Universal childcare, maternal labor supply, and family well-being. National Bureau of Economic Research, No. w11832.

Baldwin, M.L., Butler, R.J., Johnson, W.G., 2001. A hierarchical theory of occupational segregation and wage discrimination. Econ. Inq. 39 (1), 94–110.

Barasinska, N., Schäfer, D., 2013. Financial risk taking, gender and social identity—evidence from national surveys of household finance. LWS WP No. 15.

Barber, B.M., Odean, T., 2001. Boys will be boys: gender, overconfidence, and common stock investment. Q. J. Econ. 116 (1), 261–292.

Bárcena-Martín, E., Moro-Egido, A.I., 2013. Gender and poverty risk in Europe. Fem. Econ. 19 (2), 69–99.

Bardasi, E., Jenkins, S.P., 2010. The gender gap in private pensions. Bull. Econ. Res. 62, 343–363.

Barg, K., Beblo, M., 2009. Does marriage pay more than cohabitation? J. Econ. Stud. 36 (6), 552–570.

Barg, K., Beblo, M., 2012. Does sorting into specialization' explain the differences in time use between married and cohabiting couples? An empirical application for Germany. Ann. Econ. Stat. 105–106, 127–154.

Barón, J.D., Cobb-Clark, D.A., 2010. Occupational segregation and the gender wage gap in private—and public—sector employment: a distributional analysis. Econ. Rec. 86 (273), 227–246.

Barron, D., West, E., 2013. The financial costs of caring in the British labour market: is there a wage penalty for workers in caring occupations? Br. J. Ind. Relat. 51 (1), 104–123.

Barth, E., Dale-Olsen, H., 2009. Monopsonistic discrimination, worker turnover, and the gender wage gap. Labour Econ. 16 (5), 589–597.

Basu, K., 2006. Gender and say: a model of household behaviour with endogenously determined balance of power. Econ. J. 116, 558–580.
Batalova, J.A., Cohen, P., 2002. Premarital cohabitation and housework: couples in cross-national perspective. J. Marriage Fam. 64, 743–755.
Bauer, T.K., Sinning, M., 2008. An extension of the Blinder–Oaxaca decomposition to nonlinear models. Adv. Stat. Anal. 92 (2), 197–206.
Baxter, J., 2005. To marry or not to marry. Marital status and the household division of labor. J. Fam. Issues 26, 300–321.
Baxter, J., Hewitt, B., 2013. Negotiating domestic labor: women's earnings and housework time in Australia. Fem. Econ. 19 (1), 29–53.
Baxter, J., Haynes, M., Hewitt, B., 2010. Pathways into marriage: cohabitation and the domestic division of labor. J. Fam. Issues 31 (11), 1507–1529.
Bayard, K., Hellerstein, J., Neumark, D., Troske, K., 2003. New evidence on sex segregation and sex differences in wages from matched employee–employer data. J. Labor Econ. 21 (4), 887–922.
Beblo, M., Robledo, J.R., 2008. The wage gap and the leisure gap for double-earner couples. J. Popul. Econ. 21 (2), 281–304.
Beblo, M., Bender, S., Wolf, E., 2009. Establishment-level wage effects of entering motherhood. Oxf. Econ. Pap. 61 (Suppl. 1), i11–i34.
Becker, G.S., 1965. A theory of the allocation of time. Econ. J. 75 (299), 493–517.
Becker, G.S., 1971. The Economics of Discrimination, second ed. University of Chicago Press, Chicago.
Becker, G.S., 1973. A theory of marriage: part I. J. Polit. Econ. 81 (4), 813–846.
Becker, G.S., 1974a. A theory of social interactions. J. Polit. Econ. 82 (6), 1063–1093.
Becker, G.S., 1974b. A theory of marriage, part 2. J. Polit. Econ. 82 (2), S11–S26.
Becker, G.S., 1985. Human capital, effort and the sexual division of labor. J. Labor Econ. 3 (1-part 2), S33–S58.
Becker, G.S., 1991. A Treatise on the Family, second ed. Cambridge University Press, Cambridge.
Becker-Blease, J.R., Sohl, J.E., 2007. Do women-owned businesses have equal access to angel capital? J. Bus. Ventur. 22, 503–521.
Behrendt, C., 2000. Private pensions—a viable alternative? Their distributive effects in a comparative perspective. Int. Soc. Secur. Rev. 53 (3), 3–26.
Bell, L.A., 2005. Women-led firms and the gender gap in top executive jobs. IZA Discussion Papers No. 1689.
Bellante, D., Foster, A.C., 1984. Working wives and expenditure on services. J. Consum. Res. 11 (2), 700–707.
Bellante, D., Link, A.N., 1981. Are public sector workers more risk averse than private sector workers? Ind. Labor Relat. Rev. 34 (3), 408–412.
Bennett, F., 2013. Researching within-household distribution: overview, developments, debates, and methodological challenges. J. Marriage Fam. 75, 582–597.
Bennett, F., Daly, M., 2014. Poverty through a gender lens: evidence and policy review on gender and poverty. Working Paper for the Joseph Rowntree Foundation, University of Oxford, Department of Social Policy.
Ben-Porath, Y., 1982. Economics and the family—match or mismatch? A review of Becker's a treatise on the family. J. Econ. Lit. 20 (1), 52–64.
Berger, L., Waldfogel, J., 2004. Maternity leave and the employment of new mothers in the United States. J. Popul. Econ. 17 (2), 331–349.
Bergmann, B., 1974. Occupational segregation, wages, and profits when employers discriminate by race or sex. East. Econ. J. 1 (1–2), 103–110.
Bergmann, B., 1995. Becker's theory of the family: preposterous conclusions. Fem. Econ. 1 (1), 141–150.
Bergstrom, T., 1989. A fresh look at the rotten-kid theorem and other household mysteries. J. Polit. Econ. 97 (5), 1138–1159.
Bergstrom, T.C., 1996. Economics in a family way. J. Econ. Lit. 34 (4), 1903–1934.
Bertrand, M., 2010. New perspectives on gender. In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics, North Holland, vol. 4b, pp. 1543–1590 (Chapter 17).

Bertrand, M., Hallock, K.F., 2001. The gender gap in top corporate jobs. Ind. Labor Relat. Rev. 55 (1), 3–21.
Bertrand, M., Goldin, C., Katz, L.F., 2010. Dynamics of the gender gap for young professionals in the financial and corporate sectors. Am. Econ. J. 2 (3), 228–255.
Bertrand, M., Kamenika, E., Pan, J., 2013. Gender identity and relative income within households. Chicago Booth Working Paper No. 13-08.
Bettio, F., Tinios, P., Betti, G., 2013. The Gender Gap in Pensions in the EU. Publications Office of the European Union, Luxembourg.
Bianchi, S.M., Rytina, N., 1986. The decline in occupational sex segregation during the 19705: census and cps comparisons. Demography 23 (1), 79–86.
Bianchi, S.M., Sayer, L.C., Milkie, M.A., Robinson, J.P., 2012. Housework: who did, does or will do it, and how much does it matter? Soc. Forces 91 (1), 55–63.
Bielby, D.D., Bielby, W.T., 1988. She works hard for the money: household responsibilities and the allocation of work effort. Am. J. Sociol. 93, 1031–1059.
Bittman, M., Wajcman, J., 2000. The rush hour: the character of leisure time and gender equity. Soc. Forces 79 (1), 165–189.
Bittman, M., England, P., Folbre, N., Sayer, L., Matheson, G., 2003. When does gender trump money? Bargaining and time in household work. Am. J. Sociol. 109 (1), 186–214.
Bittman, M., Rice, J.M., Wajcman, J., 2004. Appliances and their impact: the ownership of domestic technology and time spent on household work. Br. J. Sociol. 55 (3), 401–423.
Black, S.E., Brainerd, E., 2004. The impact of globalization on gender discrimination. Ind. Labor Relat. Rev. 57, 540–559.
Black, S.E., Juhn, C., 2000. The rise of female professionals: women's response to rising skill demand. Am. Econ. Rev. 90 (2), 450–455.
Black, S., Strahan, P.E., 2001. The division of spoils—rent sharing and discrimination in a regulated industry. Am. Econ. Rev. 91 (4), 814–831.
Black, S.E., Spitz-Oener, A., 2010. Explaining women's success: technological change and the skill content of women's work. Rev. Econ. Stat. 92 (1), 187–194.
Black, D.A., Haviland, A.M., Sanders, S.G., Taylor, L.J., 2008. Gender wage disparities among the highly educated. J. Hum. Resour. 43 (3), 630–659.
Blackburn, M., Korenman, S., 1994. The declining marital-status earnings differential. J. Popul. Econ. 7, 249–270.
Blanchard, L., Zhao, Bo, Yinger, J., 2008. Do lenders discriminate against minority and woman entrepreneurs? J. Urban Econ. 63 (2), 467–497.
Blanchflower, D.G., Wainwright, J., 2005. An analysis of the impact of affirmative action programs on self-employment in the construction industry. National Bureau of Economic Research, Working Paper No. 11793.
Blanchflower, D.G., Levine, P.B., Zimmerman, D.J., 2003. Discrimination in the small-business credit market. Rev. Econ. Stat. 85 (4), 930–943.
Blank, R.M., 1985. An analysis of workers' choice between employment in the public and private sectors. Ind. Labor Relat. Rev. 38 (2), 211–224.
Blau, F.D., 1977. Equal Pay in the Office. Lexington Books, Lexington, MA.
Blau, D., 2003. Child care subsidy programs. In: Moffitt (Ed.), Means-Tested Transfer Programs in the United States. University of Chicago Press, pp. 443–516 (Chapter 7).
Blau, F.D., 2012. Gender, Inequality, and Wages. Oxford University Press, p. 576.
Blau, F.D., Hendricks, W.E., 1979. Occupational segregation by sex: trends and prospects. J. Hum. Resour. 14 (2), 197–210.
Blau, F.D., Kahn, L.M., 1992. The gender earnings gap: learning from international comparisons. Am. Econ. Rev. 82 (2), 533–538.
Blau, F.D., Kahn, L.M., 1996. Wage structure and gender earnings differentials: an international comparison. Economica 63, S29–S62.
Blau, F., Kahn, L.M., 1997. Swimming upstream: trends in the gender wage differential in the 1980s. J. Labor Econ. 15 (1), 1–42.

Blau, F., Kahn, L.M., 2003. Understanding international differences in the gender pay gap. J. Labor Econ. 21 (1), 106–144.
Blau, F.D., Kahn, L.M., 2006a. The gender pay gap: going, going, going, . . . but not gone. In: Blau, F.D., Brinton, M.C., Grugsky, D.B. (Eds.), The Declining Significance of Gender? Russell Sage Foundation, New York, NY, pp. 37–66 (Chapter 2).
Blau, F.D., Kahn, L.M., 2006b. The US gender pay gap in the 1990s: slowing convergence. Ind. Labor Relat. Rev. 60, 45–66.
Blau, F.D., Kahn, L.M., 2007. Changes in the labor supply behavior of married women: 1980–2000. J. Labor Econ. 25 (3), 393–438.
Blau, F.D., Kahn, L.M., 2013. Female labor supply: why is the United States falling behind? Am. Econ. Rev. 103 (3), 251–256.
Blau, F.D., Simpson, P., Anderson, D., 1998. Continuing progress? Trends in occupational segregation in the United States over the 1970s and 1980s. Fem. Econ. 4 (3), 29–71.
Blau, F.D., Brummund, P., Liu, A.Y.H., 2013. Trends in occupational segregation by gender 1970–2009: adjusting for the impact of changes in the occupational coding system. Demography 50 (2), 471–492.
Blinder, A.S., 1973. Wage discrimination: reduced form and structural estimates. J. Hum. Resour. 8 (4), 436–455.
Bloemen, H.G., Stancanelli, E.G., 2014. Market hours, household work, child care, and wage rates of partners: an empirical analysis. Rev. Econ. Househ. 12, 51–81. http://dx.doi.org/10.1007/s11150-013-9219-4.
Blood, R.O., Wolfe, D.M., 1960. Husbands and Wives: The Dynamics of Married Living. Free Press, New York, NY.
Blundell, R., Gosling, A., Ichimura, H., Meghir, C., 2007. Changes in the distribution of female and male wages accounting for employment composition using bounds. Econometrica 75 (2), 323–363.
Blundell, R., Bozio, A., Laroque, G., 2011a. Labor supply and the extensive margin. Am. Econ. Rev. 101 (3), 482–486.
Blundell, R., Bozio, A., Laroque, G., 2011b. Extensive and intensive margins of labour supply: working hours in the US, UK and France. IFS Working Papers No. 11-01.
Boden, R.J., 1999. Flexible working hours, family responsibilities, and female self-employment. Am. J. Econ. Sociol. 58 (1), 71–83.
Bonilla, R., Kiraly, F., 2013. Marriage wage premium in a search equilibrium. Labour Econ. 24, 107–115.
Bonin, H., Dohmen, T., Falk, A., Huffman, D., Sunde, U., 2007. Cross-sectional earnings risk and occupational sorting: the role of risk attitudes. Labour Econ. 14 (6), 926–937.
Bonke, J., 1992. Distribution of economic resources: implications of including household production. Rev. Income Wealth 38 (3), 281–293.
Bonke, J., 2013. Pooling of income and sharing of consumption within households. Rev. Econ. Househ. http://dx.doi.org/10.1007/s11150-013-9184-y.
Bonke, J., Browning, M., 2009. The distribution of financial well-being and income within the household. Rev. Econ. Househ. 7, 31–42.
Bonke, J., Gerstoft, F., 2007. Stress, time use and gender. Int. J. Time Use Res. 4 (1), 47–68.
Bonke, J., Uldall-Poulsen, H., 2007. Why do families actually pool their income? Evidence from Denmark. Rev. Econ. Househ. 5, 113–128.
Bonke, J., Datta-Gupta, N., Smith, N., 2005. Timing and flexibility of housework and men and women's wages. In: Hamermesh, D.S., Pfann, G.A. (Eds.), Contributions to Economic Analysis, vol. 271. Elsevier Press, Amsterdam, pp. 43–78.
Bonnet, C., Geraci, M., 2009. Correcting gender inequalities in pensions. The experience of five countries. INED Population and Societies, n°453.
Bonnet, C., Hourriez, J.M., 2012. Gender equality in pensions: what role for rights accrued as a spouse or a parent? Population (Engl. ed.) 67 (1), 123–146.
Bonnet, C., Buffeteau, S., Godefroy, P., 2006. Effects of pension reforms on gender inequality in France. Population (Engl. ed.) 61 (1), 41–70.
Bonnet, C., Keogh, A., Rapoport, B., 2013. How can we explain the gender wealth gap in France? Ined Working Paper No. 191.

Booth, A., Leigh, A., 2010. Do employers discriminate by gender? A field experiment in female-dominated occupations. Econ. Lett. 107 (2), 236–238.

Booth, A., Nolen, P., 2012a. Choosing to compete: how different are girls and boys? J. Econ. Behav. Organ. 81 (2), 542–555.

Booth, A.L., Nolen, P., 2012b. Gender differences in risk behaviour: does nurture matter? Econ. J. 122 (558), F56–F78.

Booth, A.L., Van Ours, J.C., 2008. Job satisfaction and family happiness: the part-time work puzzle. Econ. J. 118 (526), F77–F99.

Booth, A.L., van Ours, J., 2009. Hours of work and gender identity: does part-time work make the family happier? Economica 76 (301), 176–196.

Booth, A.L., van Ours, J.C., 2013. Part-time jobs: what women want? J. Popul. Econ. 26 (1), 263–283.

Booth, A.L., Francesconi, M., Frank, J., 2003. A sticky floors model of promotion, pay, and gender. Eur. Econ. Rev. 47 (2), 295–322.

Borooah, V.K., McKee, P.M., 1993. Intra-household income transfers and implications for poverty and inequality in the UK. In: Creedy, J. (Ed.), Taxation, Poverty and the Distribution of Income. Edward Elgar, London.

Boskin, M.J., 1975. Efficiency aspects of the differential tax treatment of market and household economic activity. J. Public Econ. 4 (1), 1–25.

Boskin, M.J., Sheshinski, E., 1983. Optimal tax treatment of the family: married couples. J. Public Econ. 20 (3), 281–297.

Bourguignon, F., 1984. Rationalité individuelle ou rationalité stratégique: le cas de l'offre familiale de travail. Rev. Econ. 35 (1), 147–162.

Bourguignon, F., Chiappori, P.-A., 1992. Collective models of household behavior. Eur. Econ. Rev. 36 (2), 355–364.

Bourguignon, F., Browning, M., Chiappori, P.-A., Lechene, V., 1993. Intra household allocation of consumption: a model and some evidence from French data. Ann. Econ. Stat. 29, 137–156.

Bover, O., 2010. Wealth inequality and household structure: US vs. Spain. Rev. Income Wealth 56 (2), 259–290.

Boye, K., 2009. Relatively different? How do gender differences in well-being depend on paid and unpaid work in Europe? Soc. Indic. Res. 93 (3), 509–525.

Brady, D., Kall, D., 2008. Nearly universal, but somewhat distinct: the feminization of poverty in affluent Western democracies, 1969–2000. Soc. Sci. Res. 37 (3), 976–1007.

Breen, R., Cooke, L.P., 2005. The persistence of the gendered division of domestic labour. Eur. Sociol. Rev. 21 (1), 43–57.

Breunig, R., Cobb-Clark, D.A., Gong, X., Venn, D., 2007. Disagreement in Australian partners' reports of financial difficulty. Rev. Econ. Household 5 (1), 59–82.

Brines, J., 1994. Economic dependency, gender and the division of labor at home. Am. J. Sociol. 100 (3), 652–688.

Brown, C., Corcoran, M., 1997. Sex-based differences in school content and the male/female wage gap. J. Labor Econ. 15, 431–465.

Browning, M., Chiappori, P.-A., 1998. Efficient intra-household allocation: a characterisation and tests. Econometrica 66 (6), 1241–1278.

Browning, M., Gørtz, M., 2012. Spending time and money within the household. Scand. J. Econ. 114 (3), 681–704.

Browning, M., Meghir, C., 1991. The effects of male and female labor supply on commodity demands. Econometrica 59 (4), 925–951.

Browning, M., Bourguignon, F., Chiappori, P.-A., Lechene, V., 1994. Income and outcomes: a structural model of intrahousehold allocation. J. Polit. Econ. 102 (6), 1067–1096.

Browning, M., Chiappori, P.-A., Lewbel, A., 2006a. Estimating consumption economies of scale, adult equivalence scale, and household bargaining power. Oxford Department of Economics, Discussion Paper No. 289.

Browning, M., Chiappori, P.-A., Lechene, V., 2006b. Collective and unitary models: a clarification. Rev. Econ. Househ. 4 (1), 5–14.

Browning, M., Chiappori, P.A., Weiss, Y., 2011. Family Economics. CUP, Cambridge.
Bryan, M.L., Sevilla-Sanz, A., 2011. Does housework lower wages? Evidence for Britain. Oxf. Econ. Pap. 63 (1), 187–210.
Bryant, W.K., Zick, C.D., 1996. Are we investing less in the next generation? Historical trends in time spent caring for children. J. Fam. Econ. Iss. 17, 365–392.
Budig, M.J., 2006a. Intersections on the road to self-employment: gender, family and occupational class. Soc. Forces 84 (4), 2223–2239.
Budig, M.J., 2006b. Gender, self-employment, and earnings—the interlocking structures of family and professional status. Gend. Soc. 20 (6), 725–753.
Budig, M.T., England, P., 2001. The wage penalty for motherhood. Am. Sociol. Rev. 66 (2), 204–225.
Burda, M.C., Hamermesh, D.S., Weil, P., 2006. The distribution of total work in the EU and US. IZA Discussion Papers No. 2270.
Burda, M., Hamermesh, D.S., Weil, P., 2013. Total work and gender: facts and possible explanations. J. Popul. Econ. 26 (1), 239–261.
Burdett, K., Mortensen, D.T., 1998. Wage differentials, employer size, and unemployment. Int. Econ. Rev. 39 (2), 257–273.
Burgoyne, C.B., 1990. Money in marriage: how patterns of allocation both reflect and conceal power. Sociol. Rev. 38 (4), 634–665.
Burgoyne, C.B., Morison, V., 1997. Money in remarriage: keeping things simple—and separate. Sociol. Rev. 45 (3), 363–395.
Burgoyne, C.B., Reibstein, J., Edmunds, A., Dolman, V., 2007. Money management systems in early marriage: factors influencing change and stability. J. Econ. Psychol. 28, 214–228.
Canberra Group, 2001. Expert Group on Household Income Statistics: Final Report and Recommendations. Canberra Group, Ottawa.
Canberra Group, 2011. Handbook on Household Income Statistics, second ed. United Nations Economic Commission for Europe, Geneva.
Cannings, K., Montmarquette, C., 1990. Managerial momentum: a simultaneous model of the career progress of male and female managers. Indus. Lab. Rel. Rev 44 (2), 212.
Cantillon, R., 1952/1755. Essai sur la Nature du Commerce en General. Ined, Paris.
Cantillon, S., 2013. Measuring differences in living standards within households. J. Marriage Fam. 75, 598–610.
Cantillon, S., Nolan, B., 1998. Are married women more deprived than their husbands? J. Soc. Policy 27, 151–171.
Carr, D., 1996. Two paths to self-employment? Women's and men's self-employment in the United States, 1980. Work. Occup. 23 (1), 26–53.
Carrell, S.E., Page, M.E., West, J.E., 2010. Sex and science: how professor gender perpetuates the gender gap. Q. J. Econ. 125 (3), 1101–1144.
Carrington, W.J., Troske, K.R., 1995. Gender segregation in small firms. J. Hum. Res. 30 (3), 503–533.
Carter, M.R., Katz, E.G., 1993. Separate spheres and the conjugal contract: understanding the impact of gender-biased development. In: Haddad, et al. (Eds.), Intrahousehold Resource Allocation in Developing Countries: Methods, Models, and Policy, International Food Policy Research Institute, pp. 95–111.
Carter, D.A., D'Souza, F., Simkins, B.J., Simpson, W.G., 2007. The diversity of corporate board committees and financial performance. Oklahoma State University, Working Paper, 89–154.
Caucutt, E.M., Guner, N., Knowles, J., 2002. Why do women wait? Matching, wage inequality, and the incentives for fertility delay. Rev. Econ. Dyn. 5 (4), 815–855.
Chadeau, A., 1992. What is households' non-market production worth? OECD Econ. Stud. 18, 85–103.
Chand, H., Gan, L., 2003. The effects of bracketing in wealth estimation. Rev. Income Wealth 49 (2), 273–287.
Chang, C.F., England, P., 2011. Gender inequality in earnings in industrialized East Asia. Soc. Sci. Res. 40 (1), 1–14.
Chatterji, M., Mumford, K., Smith, P.N., 2011. The public–private sector gender wage differential in Britain: evidence from matched employee–workplace data. Appl. Econ. 43 (26), 3819–3833.

Chatterji, A.K., Chay, K.Y., Fairlie, R.W., 2013. The impact of city contracting set-asides on black self-employment and employment. National Bureau of Economic Research No. w18884.
Chen, Z., Woolley, F., 2001. A Cournot–Nash model of family decision making. Econ. J. 111 (474), 722–748.
Chernozhukov, V., Fernández-Val, I., Melly, B., 2013. Inference on counterfactual distributions. Econometrica 81 (6), 2205–2268.
Chiappori, P.-A., 1988. Rational household labor supply. Econometrica 56 (1), 63–90.
Chiappori, P.-A., 1992. Collective labor supply and welfare. J. Polit. Econ. 100 (3), 437–467.
Chiappori, P.-A., 1997. Introducing household production in collective models of labor supply. J. Polit. Econ. 105, 191–209.
Chiappori, P.-A., Meghir, C., 2014. Intrahousehold inequality. In: Atkinson, Bourguignon (Eds.), Handbook of Income Distribution, This volume (Chapter 16).
Christofides, L.N., Polycarpou, A., Vrachimis, K., 2013. Gender wage gaps, 'sticky floors' and 'glass ceilings' in Europe. Labour Econ. 21, 86–102.
Chun, H., Lee, I., 2001. Why do married men earn more: productivity or marriage selection? Econ. Inq. 39 (2), 307–319.
Cigno, A., 2008. A gender-neutral approach to gender issues. In: Dex, S., Scott, J. (Eds.), Frontiers in the Economics of Gender. Routledge, London, pp. 45–56.
Clain, S.H., 2000. Gender differences in full-time self-employment. J. Econ. Bus. 52 (6), 499–513.
Clark, A., Postel-Vinay, F., 2009. Job security and job protection. Oxf. Econ. Pap. 61 (2), 207–239.
Coate, S., Loury, G.C., 1993. Will affirmative-action policies eliminate negative stereotypes? Am. Econ. Rev. 83 (5), 1220.
Cohen, P.N., Huffman, M.L., 2007. Working for the woman? Female managers and the gender wage gap. Am. Sociol. Rev. 72 (5), 681–704.
Cole, R.A., Mehran, H., 2009. Gender and the availability of credit to privately held firms: evidence from the surveys of small business finances. Federal Reserve Bank of New York Staff.
Coleman, M.T., 1988. The division of household labor suggestions for future empirical consideration and theoretical development. J. Fam. Issues 9 (1), 132–148.
Coltrane, S., 2000. Research on household labor: modeling and measuring the social embeddedness of routine family work. J. Marriage Fam. 62 (4), 1208–1233.
Conley, D., Ryvicker, M., 2005. The price of female headship: gender, inheritance, and wealth accumulation in the United States. J. Income Distrib. 13, 41–56.
Connelly, R., 1992. Self-employment and providing child care. Demography 29 (1), 17–29.
Connelly, R., Kimmel, J., 2009. Spousal influences on parents' non-market time choices. Rev. Econ. Househ. 7 (4), 361–394.
Connolly, S., Gregory, M., 2008. Moving down: women's part-time work and occupational change in Britain 1991–2001. Econ. J. 118 (526), F52–F76.
Cornwell, C., Rupert, P., 1997. Unobservable individual effects, marriage and the earnings of young men. Econ. Inq. 35 (2), 285–294.
Cotton, J., 1988. On the decomposition of wage differentials. Rev. Econ. Stat. 70 (2), 236–243.
Couprie, H., 2007. Time allocation within the family: welfare implications of life in a couple. Econ. J. 117 (516), 287–305.
Couprie, H., Ferrant, G., 2014. Welfare comparisons, economies of scale and equivalence scale in time use. Ann. Econ. Stat. forthcoming.
Coverman, S., 1985. Explaining husbands' participation in domestic labor. Sociol. Q. 26, 81–97.
Cowell, F., Karagiannaki, E., McKnight, A., 2012. Mapping and measuring the distribution of household wealth: a cross-country analysis. Luxembourg Wealth Study WP Series 12.
Cox, D., 2003. Private transfers within the family: mothers, fathers, sons and daughters. In: Munnell, A., Sundén, A. (Eds.), Death and Dollars: The Role of Gifts and Bequests in America. Brookings Institution Press, Washington, DC, pp. 168–210.
Craig, L., 2006. Children and the revolution. A time-diary analysis of the impact of motherhood on daily workload. J. Sociol. 42 (2), 125–143.

Craig, L., 2007. Is there really a second shift, and if so, who does it? A time-diary investigation. Fem. Rev. 86 (1), 149–170.
Craig, L., Bittman, M., 2008. The incremental time costs of children: an analysis of children's impact on adult time use in Australia. Fem. Econ. 14 (2), 59–88.
Craig, L., Mullan, K., 2010. Parenthood, gender and work-family time in the United States, Australia, Italy, France, and Denmark. J. Marriage Fam. 72 (5), 1344–1361.
Craig, L., Siminski, P., 2011. If men do more housework, do their wives have more babies? Soc. Indic. Res. 101 (2), 255–258.
Craig, L., Powell, A., Cortis, N., 2012. Self-employment, work-family time and the gender division of labour. Work Employ. Soc. 26 (5), 716–734.
Croson, R., Gneezy, U., 2009. Gender differences in preferences. J. Econ. Lit. 47 (20), 448–474.
D'Addio, A.C., 2013. Pension entitlements of women with children: the role of credits within pension systems in OECD and EU countries. In: Holzmann, R., Palmer, E.E., Robalino, D. (Eds.), Pension Reform: Issues and Prospects for Non-Financial Defined Contribution (NDC) Schemes, vol. 2. Gender, Politics and Financial Stability. World Bank Publications, pp. 75–110.
Datta Gupta, N., Smith, N., 2002. Children and career interruptions: the family gap in Denmark. Economica 69 (276), 609–629.
Datta Gupta, N., Stratton, L.S., 2010. Examining the impact of alternative power measures on individual time use in American and Danish couple households. Rev. Econ. Househ. 8 (3), 325–343.
Datta Gupta, N.D., Oaxaca, R.L., Smith, N., 2006. Swimming upstream, floating downstream: comparing women's relative wage progress in the United States and Denmark. Ind. Labor Relat. Rev. 59 (2), 243–266.
Datta Gupta, N., Smith, N., Stratton, L.S., 2007. Is marriage poisonous? Are relationships taxing? An analysis of the male marital wage differential in Denmark. South. Econ. J. 74, 412–433.
Datta Gupta, N., Poulsen, A., Villeval, M.C., 2013. Gender matching and competitiveness: experimental evidence. Econ. Inq. 51 (1), 816–835.
Datta Gupta, N., Smith, N., Verner, M., 2008. The impact of Nordic countries family friendly policies on employment, wages and children. Rev. Econ. Househ. 6, 65–89.
Davies, H., Joshi, H., 1994. Sex, sharing and the distribution of income. J. Soc. Policy 23 (3), 301–340.
Davies, R., Pierre, G., 2005. The family gap in pay in Europe: a cross-country study. Labour Econ. 12 (4), 469–486.
Davies, J.B., Shorrocks, A.F., 2000. The distribution of wealth. In: Atkinson, A.B., Bourguignon, F. (Eds.), Handbook of Income Distribution. In: vol. 1. Elsevier, pp. 605–675.
Davis, J.B., 2006. Social identity strategies in recent economics. J. Econ. Methodol. 13 (3), 371–390.
Davis, S.N., Greenstein, T.N., 2004. Cross-national variations in the division of household labor. J. Marriage Fam. 66 (5), 1260–1271.
Davis, S.N., Wills, J.B., 2013. Theoretical explanations amid social change. A content analysis of housework research (1975–2012). J. Fam. Issues. 35 (6), 808–824.
Davis, B., Handa, S., Ruiz, M., Stampini, M., Winters, P., 2002. Conditionality and the impact of program design on household welfare: comparing two diverse cash transfer programs in rural Mexico. ESA/FAO WP 07-2002.
Davis, S., Greenstein, T., Marks, J., 2007. Effects of union type on division of household labor: do cohabiting men really perform more housework? J. Fam. Issues 28, 1246–1272.
de Castro, F., Salto, M., Steiner, H., 2013. The gap between public and private wages: new evidence for the EU (No. 508). Directorate General Economic and Monetary Affairs (DG ECFIN), European Commission.
De Henau, J., Himmelweit, S., Santos, C., Soobedar, Z., 2013. Comparing welfare regimes by their effects on intra-household inequalities. In: Garces, J., Monsonis-Paya, I. (Eds.), Sustainability and Transformation in European Social Policy. Peter Lang, Oxford, pp. 117–146.
De Laat, J., Sevilla-Sanz, A., 2011. The fertility and women's labor force participation puzzle in OECD countries: the role of men's home production. Fem. Econ. 17 (2), 87–119.
De Ruijter, E., Treas, J.K., Cohen, P.N., 2005. Outsourcing the gender factory: living arrangements and service expenditures on female and male tasks. Soc. Forces 84 (1), 305–322.

Decancq, K., Fleurbaey, M., Schokkaert, E., 2014. Inequality, income, and well-being. In: Atkinson, Bourguignon (Eds.), Handbook of Income Distribution, This volume (Chapter 2).
Deere, C.D., Doss, C.R., 2006. The gender asset gap: what do we know and why does it matter? Fem. Econ. 12 (1–2), 1–50.
Del Boca, D., 2002. The effect of child care and part time opportunities on participation and fertility decisions in Italy. J. Popul. Econ. 15 (3), 549–573.
Del Boca, D., Pasqua, S., 2003. Employment patterns of husbands and wives and family income distribution in Italy (1977–98). Rev. Income Wealth 49 (2), 221–245.
Del Boca, D., Pasqua, S., Pronzato, C., 2009. Motherhood and market work decisions in institutional context: a European perspective. Oxf. Econ. Pap. 61 (Suppl. 1), i147–i171.
Del Río, C., Gradín, C., Cantó, O., 2011. The measurement of gender wage discrimination: the distributional approach revisited. J. Econ. Inequal. 9 (1), 57–86.
Denton, M., Boos, L., 2007. The gender wealth gap: structural and material constraints and implications for later life. J. Women Aging 19 (3–4), 105–120.
Depalo, D., Pereira, M.C., Eugène, B., Papapetrou, E., Perez, J.J., Reiss, L., Roter, M., 2011. The public sector pay gap in a selection of Euro area countries. Working Paper Series No. 1406. European Central Bank.
Devine, T.J., 1994. Changes in wage-and-salary returns to skill and the recent rise in female self-employment. Am. Econ. Rev. 84 (2), 108–113.
Dex, S., 2010. Can state policies produce equality in housework. In: Treas, J., Drobnic, S. (Eds.), Dividing the Domestic: Men, Women, and Household Work in Cross-National Perspective. Stanford University Press, pp. 79–104.
Dina, Shatnawi, Ronald, Oaxaca, Michael, Ransom, 2012. Movin' on up: hierarchical occupational segmentation and gender wage gaps. IZA DP No. 7011.
DiNardo, J., Fortin, N., Lemieux, T., 1996. Labour market institutions and the distribution of wages, 1973–1992: a semi-parametric approach. Econometrica 64 (5), 1001–1044.
Dohmen, T., Falk, A., 2011. Performance pay and multidimensional sorting: productivity, preferences, and gender. Am. Econ. Rev. 101 (2), 556–590.
Dolado, J.J., Felgueroso, F., Jimeno, J.F., 2003. Where do women work? Analysing patterns in occupational segregation by gender. Ann. Econ. Stat. 71–72, 293–315.
Domínguez-Folgueras, M., 2012. Is cohabitation more egalitarian? The division of household labor in five European countries. J. Fam. Issues 34 (12), 1623–1646.
Donni, O., Chiappori, P.-A., 2011. Nonunitary models of household behavior: a survey of the literature. In: Molina, J.A. (Ed.), Household Economics Behaviours, International Series on Consumer Science. Springer, New York, NY, pp. 1–40.
Dougherty, C., 2006. The marriage earnings premium as a distributed fixed effect. J. Hum. Resour. 41 (2), 433–443.
Dreber, A., von Essen, E., Ranehill, E., 2011. Outrunning the gender gap—boys and girls compete equally. Exp. Econ. 14 (4), 567–582.
Dribe, M., Stanfors, M., 2009. Does parenthood strengthen a traditional household division of labor? Evidence from Sweden. J. Marriage Fam. 71 (1), 33–45.
Drolet, M., Mumford, K., 2012. The gender pay gap for private—sector employees in Canada and Britain. Br. J. Ind. Relat. 50 (3), 529–553.
Du Rietz, A., Henrekson, M., 2000. Testing the female underperformance hypothesis. Small Bus. Econ. 14 (1), 1–10.
Duflo, E., 2003. Grandmothers and granddaughters: old-age pensions and intrahousehold allocation in South Africa. World Bank Econ. Rev. 17, 1–25.
Duflo, E., Saez, E., 2003. The role of information and social interactions in retirement plan decisions: evidence from a randomized experiment. Q. J. Econ. 118, 815–842.
Duguet, E., Petit, P., Petit, P., 2005. Hiring discrimination in the French financial sector: an econometric analysis on field experiment data. Ann. Econ. Stat. 78, 79–102.
Dustmann, C., Schonberg, U., 2012. Expansions in maternity leave coverage and children's long-term outcomes. Am. Econ. J. 4 (3), 190–224.

Dwyer, P.D., Gilkenson, J.H., List, J.A., 2002. Gender differences in revealed risk taking, evidence from mutual fund investors. Econ. Lett. 76 (2), 151–158.
Eberts, R.W., Stone, J.A., 1985. Male female differences in promotions: EEO in public education. J. Hum. Resour. 20 (4), 504–521.
Eckel, C.C., Grossman, P.J., 2008. Sex and risk: experimental evidence. In: Plott, C., Smith, V. (Eds.), Handbook of Experimental Economics Results. In: vol. 1. Elsevier, New York, NY.
Edlund, L., Kopczuk, W., 2009. Women, wealth, and mobility. Am. Econ. Rev. 99 (1), 146–178.
Ekberg, J., Eriksson, R., Friebel, G., 2013. Parental leave—a policy evaluation of the Swedish "Daddy-Month" reform. J. Public Econ. 97, 131–143.
Ekelund, J., Johansson, E., Järvelin, M.R., Lichtermann, D., 2005. Self-employment and risk aversion—evidence from psychological test data. Labour Econ. 12 (5), 649–659.
Elder, H.W., Rudolph, P.M., 2003. Who makes the financial decisions in the households of older Americans? Finan. Serv. Rev. 12 (4), 293–308.
Elizabeth, V., 2001. Managing money, managing coupledom: a critical examination of cohabitants' money management practices. Sociol. Rev. 49 (3), 389–411.
England, P., 1982. The failure of human capital theory to explain occupational sex segregation. J. Hum. Resour. 17, 358–370.
England, P., 2010. The gender revolution: uneven and stalled. Gend. Soc. 24, 149–166.
England, P., 2011. Missing the big picture and making much ado about almost nothing: recent scholarship on gender and household work. J. Fam. Theory Rev. 3 (1), 23–26.
Equal Opportunities Commission, 2005. Britain's Hidden Brain Drain—Final Report. Equal Opportunities Commission, Manchester.
Erhardt, N.L., Werbel, J.D., Shrader, C.B., 2003. Board of director diversity and firm financial performance. Corp. Gov. 11 (2), 102–111.
Esping-Andersen, G., 1990. The three worlds of welfare capitalism. Princeton University Press.
Esping-Andersen, G., 1999. Social Foundations of Postindustrial Economics. Oxford University Press, Oxford.
Esping-Andersen, G., 2009. The Incomplete Revolution: Adapting Welfare States to Women's New Roles. Polity Press, Cambridge.
Esping-Andersen, G., Boertien, D., Bonke, J., Gracia, P., 2013. Couple specialization in multiple equilibria. Eur. Sociol. Rev. 29 (6), 373–385.
Eurostat, 2003. Household Production and Consumption. Proposal for a Methodology of Household Satellite Accounts. Office for Official Publications of the European Communities, Luxembourg.
Eurostat, 2008. Harmonised European Time Use Surveys. Eurostat Methodologies and Working Papers. Office for Official Publications of the European Communities, Luxembourg.
Eurostat, 2010. EU-Silc 2010 module "Intra-household allocation of resources", guidelines. http://epp.eurostat.ec.europa.eu/portal/page/portal/income_social_inclusion_living_conditions/documents/tab7/2010%20Module%20-%20Guidelines%20for%20Website.pdf.
Eurostat, 2012. Intra-household sharing of resources, statistics explained (2013/12/3). http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Intra-household_sharing_of_resources#Publications.
Even, W.E., Macpherson, D.A., 2004. When will the gender gap in retirement income narrow? South. Econ. J. 71 (1), 182–200.
Evertsson, M., Nermo, M., 2004. Dependence within families and the division of labor: comparing Sweden and the United States. J. Marriage Fam. 66 (5), 1272–1286.
Evertsson, M., Nermo, M., 2007. Changing resources and the division of housework: a longitudinal study of Swedish couples. Eur. Sociol. Rev. 23 (4), 455–470.
Fairlie, Robert W., 1999. The absence of the African-American owned business: an analysis of the dynamics of self-employment. J. Labor Econ. 17 (1), 80–108.
Fairlie, R.W., 2005. An extension of the Blinder–Oaxaca decomposition technique to logit and probit models. J. Econ. Soc. Meas. 30 (4), 305–316.
Fairlie, R.W., Robb, A.M., 2009. Gender differences in business performance: evidence from the characteristics of business owners survey. Small Bus. Econ. 33, 375–395.
Felfe, C., 2012. The motherhood wage gap: what about job amenities? Labour Econ. 19 (1), 59–67.

Figari, F., 2011. From housewives to independent earners: can the tax system help Italian women to work? ISER Working Paper Series 2011-15.
Figari, F., Immervoll, H., Levy, H., Sutherland, H., 2007. Inequalities within couples: market incomes and the role of taxes and benefits in Europe. EUROMOD Working Paper Series EM6/07.
Filer, R.K., 1985. Male–female wage differences: the importance of compensating differentials. Ind. Labor Relat. Rev. 38 (3), 426–437.
Findlay, J., Wright, R., 1996. Gender, poverty and the intra-household distribution of resources. Rev. Income Wealth 42 (3), 335–351.
Firpo, S., Fortin, N.M., Lemieux, T., 2009. Unconditional quantile regressions. Econometrica 77 (3), 953–973.
Fleurbaey, M., 2009. Beyond GDP: the quest for a measure of social welfare. J. Econ. Lit. 47 (4), 1029–1075.
Folbre, N., 1986. Hearts and spades: paradigms of household economics. World Dev. 14 (2), 245–255.
Folbre, N., 1997. Gender coalitions: extrafamily influences on intrafamily inequality. In: Haddad, L. et al., (Ed.), Intrahousehold Resource Allocation in Developing Countries: Methods, Models and Policy. Johns Hopkins University Press, Baltimore, MD.
Folbre, N., 2004. A theory of the misallocation of time. In: Folbre, N., Bittman, M. (Eds.), Family Time: The Social Organization of Care. In: vol. 2. Routledge, London, pp. 7–24.
Folbre, N., 2009. Inequality and time use in the household. In: Salverda, W., Nolan, B., Smeeding, T.M. (Eds.), The Oxford Handbook of Economic Inequality. Oxford University Press, Oxford, pp. 342–363.
Folbre, N., Shaw, L., Stark, A., 2005. A special issue on gender and aging. Fem. Econ. 11 (2), 3–5.
Folbre, N., Gornick, J.C., Connolly, H., Munzi, T., 2013. Women's employment, unpaid work and economic inequality. In: Gornick, J.C., Jäntii, M. (Eds.), Income Inequality—Economic Disparities and the Middle Class in Affluent Countries. Stanford University Press, Stanford, CA, pp. 234–260 (Chapter 8).
Fonseca, R., Mullen, K.J., Zamarro, G., Zissimopoulos, J., 2012. What explains the gender gap in financial literacy? The role of household decision making. J. Consum. Aff. 46 (1), 90–106.
Fortin, N., 2005. Gender role attitudes and women's labour market outcomes across OECD countries. Oxf. Rev. Econ. Policy 21 (3), 416–438.
Fortin, N.M., 2008. The gender wage gap among young adults in the United States. The importance of money versus people. J. Hum. Resour. 43 (4), 884–918.
Fortin, N.M., Lemieux, T., 2000. Are women's wage gains men's losses? A distributional test. Am. Econ. Rev. 90, 456–460.
Fortin, N., Lemieux, T., Firpo, S., 2011. Decomposition methods in economics. In: Ashenfelter, O., Card, D.E. (Eds.), Handbook of Labor Economics, vol. 4. North Holland, pp. 1–102.
Francois, P., 1998. Gender discrimination without gender difference: theory and policy responses. J. Public Econ. 68 (1), 1–32.
Frazis, H., Stewart, J., 2011. How does household production affect measured income inequality? J. Popul. Econ. 24 (1), 3–22.
Freeman, R.B., Schettkat, R., 2005. Marketization of household production and the EU–US gap in work. Econ. Policy 20 (41), 6–50.
Frick, J.R., Grabka, M.M., Sierminska, E.M., 2007. Representative wealth data for Germany from the German SOEP: the impact of methodological decisions around imputation and the choice of the aggregation unit. DIW Discussion Paper 672.
Fritzell, J., 1999. Incorporating gender inequality into income distribution research. Int. J. Soc. Welf. 8, 56–66.
Förster, M.F., Mira d'Ercole, M., 2009. The OECD approach to measuring income distribution and poverty: strengths, limits and statistical issues. In: OECD Conference "Measuring Poverty, Income Inequality and Social Exclusion", Paris, 16–17, March.
Fryer Jr., R.G., Loury, G.C., 2005. Affirmative action and its mythology. NBER Working Paper No. w11464.
Furdas, M., Kohn, K., 2010. What's the difference?! Gender, personality, and the propensity to start a business. IZA Discussion Paper No. 4778.

Fuwa, M., 2004. Macro-level gender inequality and the division of household labor in 22 countries. Am. Sociol. Rev. 69 (6), 751–767.

Fuwa, M., Cohen, P.N., 2007. Housework and social policies. Soc. Sci. Res. 36, 512–530.

Gangl, M., Ziefle, A., 2009. Motherhood, labor force behavior and women's careers: an empirical assessment of the wage penalty for motherhood in Britain, Germany and the United-States. Demography 46 (2), 341–369.

Gardeazabal, J., Ugidos, A., 2004. More on identification in detailed wage decompositions. Rev. Econ. Stat. 86 (4), 1034–1036.

Garrat, R., Johnson, N., Weinberger, C., 2013. The state street mile: age and gender differences in competition-aversion in the field. Econ. Inq. 51 (1), 806–815.

Gayle, G.L., Golan, L., Miller, R.A., 2012. Gender differences in executive compensation and job mobility. J. Labor Econ. 30 (4), 829–872.

Geissler, H., 2012. New regulations aim to fight bogus self-employment. European Industrial Observatory Online.

Geist, C., 2005. The welfare state and the home: regime differences in the domestic division of labour. Eur. Sociol. Rev. 21 (1), 23–41.

Geist, C., 2010. Men's and women's reports about housework. In: Treas, J., Drobnic, S. (Eds.), Dividing the Domestic: Men, Women, and Household Work in Cross-National Perspective. Stanford UP, Stanford, Ca., pp. 217–240.

Geist, C., Cohen, P.N., 2011. Headed toward equality? Housework change in comparative perspective. J. Marriage Fam. 73 (4), 832–844.

Georgellis, Y., Wall, H.J., 2004. Gender differences in self-employment. Working Paper Series 1999-008C. Federal Reserve Bank of St Louis.

Gershuny, J., Robinson, J.P., 1988. Historical changes in the household division of labor. Demography 25 (4), 537–552.

Gershuny, J., Bittman, M., Brice, J., 2005. Exit, voice, and suffering: do couples adapt to changing employment patterns? J. Marriage Fam. 67, 656–665.

Gibson, J., Le, T., Scobie, G., 2006. Household bargaining over wealth and the adequacy of women's retirement incomes in New Zealand. Fem. Econ. 12 (1–2), 221–246.

Gicheva, D., 2013. Working long hours and early career outcomes in the high-end labor market. J. Labor Econ. 31 (4), 785–824.

Ginn, J., 2001. Risk of social exclusion in later life: how well do the pension systems of Britain and the US accommodate women's paid and unpaid work? Int. J. Sociol. Soc. Policy 21 (4–6), 212–244.

Ginn, Jay, Street, Debra, Arber, Sara, 2001. Women, Work, and Pensions: International Issues and Prospects. Open University Press, Philadelphia, PA.

Ginther, D.K., Zavodny, M., 2001. Is the male marriage premium due to selection? The effect of shotgun weddings on the return to marriage. J. Popul. Econ. 14 (2), 313–328.

Gneezy, U., Rustichini, A., 2004. Gender and competition at a young age. Am. Econ. Rev. 94 (2), 377–381.

Gneezy, U., Niederle, M., Rustichini, A., 2003. Performance in competitive environments: gender differences. Q. J. Econ. 118, 1049–1074.

Gneezy, U., Leonard, K.L., List, J.A., 2009. Gender differences in competition: evidence from a matrilineal and a patriarchal society. Econometrica 77 (5), 1637–1664.

Gobillon, L., Meurs, D., Roux, S., 2015. Estimating gender differences in access to jobs. J. Labor Econ. 33 (2). forthcoming.

Goda, G., Shoven, J., Slavov, S., 2007. Social security and the timing of divorce. NBER Working Paper No. 13382.

Goldin, C., 2006. The quiet revolution that transformed women's employment, education, and family. Am. Econ. Rev. 96 (2), 1–21.

Goldin, C., 2013. A pollution theory of discrimination: male and female differences in occupations and earnings. NBER Working Paper No. 8985.

Goldin, C., 2014. A grand gender convergence: its last chapter. Am. Econ. Rev. 104 (4), 1–30.

Goldin, C., Katz, L.F., 2002. The power of the pill: oral contraceptives and women's career and marriage decisions. J. Polit. Econ. 110 (4), 730–770.

Goldin, C., Katz, L.F., 2008. Transitions: career and family life cycles of the educational elite. Am. Econ. Rev. 98 (2), 363–369.
Goldin, C., Rouse, C., 2000. Orchestrating impartiality: the impact of 'blind' auditions on female musicians. Am. Econ. Rev. 90, 715–741.
Goldin, C., Katz, L.F., Kuziemko, I., 2006. The homecoming of American College Women: the reversal of the college gender gap. J. Econ. Perspect. 20 (4), 133.
Gornick, J.C., Jäntti, M., 2010. Women, poverty, and social policy regimes: a cross-national analysis. LIS Working Paper 534.
Gornick, J.C., Meyers, M.K., 2003. Welfare regimes in relation to paid work and care. Adv. Life Course Res. 8, 45–67.
Gornick, J.C., Sierminska, E., Smeeding, T.M., 2009. The income and wealth packages of older women in cross-national perspective. J. Gerontol. B Psychol. Sci. Soc. Sci. 64 (3), 402–414.
Gottschalk, P., Mayer, S., 2002. Changes in home production and trends in economic inequality. In: Cohen, D., Piketty, T., Saint-Paul, G. (Eds.), The New Economics of Rising Inequalities. Oxford University Press, New York, NY, pp. 265–284.
Grabka, M.M., Marcus, J., Sierminska, E., 2013. Wealth distribution within couples. Rev. Econ. Househ. online first, pp. 1–28.
Gray, J.S., 1997. The fall in men's return to marriage: declining productivity effects or changing selection? J. Hum. Res. 32 (3), 481–504.
Greenstein, T.N., 2000. Economic dependence, gender and the division of labor at home: a replication and extension. J. Marriage Fam. 62 (2), 322–335.
Gregory, M., 2009. Gender and economic inequality. In: Salverda, W., Nolan, B., Smeeding, T.M. (Eds.), The Oxford Handbook of Economic Inequality. Oxford University Press, Oxford, pp. 284–312.
Gregory, R.G., Borland, J., 1999. Recent developments in public sector labour markets. In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics, vol. 3C. North-Holland, Amsterdam, pp. 3573–3630.
Grimshaw, D., 2011. What do we know about low-wage work and low-wage workers? Analysing the definitions, patterns, causes and consequences in international perspective. ILO, Conditions of Work and Employment Series No. 2828.
Gronau, J., 1973. The intrafamily allocation of time: the value of the housewives' time. Am. Econ. Rev. 63 (4), 634–651.
Gronau, J., 1977. Leisure, home production, and work—the theory of the allocation of time revisited. J. Polit. Econ. 85 (6), 1099–1123.
Gronau, R., 1980. Home production—a forgotten industry. Rev. Econ. Stat. 62 (3), 408–416.
Gronau, R., Hamermesh, D.S., 2006. Time vs. goods: the value of measuring household production technologies. Rev. Income Wealth 52 (1), 1–16.
Groshen, E.L., 1991. The structure of the female/male wage differential: is it who you are, what you do, or where you work? J. Hum. Resour. 16 (3), 457–472.
Grossbard-Shechtman, A., 1974. A theory of allocation of time in markets for labour and marriage. Econ. J. 94, 863–882.
Grossbard-Shechtman, S., 2003. A consumer theory with competitive markets for work in marriage. J. Socio-Econ. 31 (6), 609–645.
Guner, N., Kaygusuz, R., Ventura, G., 2012a. Taxation and household labour supply. Rev. Econ. Stud.. .
Guner, N., Kaygusuz, R., Ventura, G., 2012b. Taxing women: a macroeconomic analysis. J. Monet. Econ. 59 (1), 111–128.
Gupta, S., 2007. Autonomy, dependence or display? The relationship between married women's earnings and housework. J. Marriage Fam. 69 (2), 399–417.
Gupta, S., Ash, M., 2008. Whose money, whose time? A nonparametric approach to modeling time spent on housework in the United States. Fem. Econ. 14 (1), 93–120.
Gupta, V.K., Turban, D.B., Wasti, S.A., Sikdar, A., 2009. The role of gender stereotypes in perceptions of entrepreneurs and intentions to become an entrepreneur. Enterp. Theory Pract. 33 (2), 397–417.
Gutierrez-Domenech, M., 2005. Employment after motherhood: a European comparison. Labour Econ. 12 (1), 99–123.

Haan, P., Wrohlich, K., 2011. Can child care policy encourage employment and fertility? Evidence from a structural model. Labour Econ. 18 (4), 498–512.
Haas, L., Hwang, C.P., 2008. The impact of taking parental leave on fathers' participation in childcare and relationships with children: lessons from Sweden. Commun. Work Fam. 11 (1), 85–104.
Haddad, L., Kanbur, R., 1990. How serious is the neglect of intra-household inequality? Econ. J. 100 (402), 866–881.
Hakim, C., 1998. Developing a sociology for the 21st century. Br. J. Sociol. 49 (1), 137–143.
Hakim, C., 2000. Work-Lifestyle Choices in the 21st Century: Preference Theory. Oxford University Press, Oxford.
Hallberg, D., Klevmarken, A., 2003. Time for children: a study of parent's time allocation. J. Popul. Econ. 16 (2), 205–226.
Halleröd, B., 2005. Sharing of housework and money among Swedish couples: do they behave rationally? Eur. Sociol. Rev. 21 (3), 273–288.
Hamermesh, D.S., Lee, J., 2007. Stressed out on four continents: time crunch or yuppie kvetch? Rev. Econ. Stat. 89 (2), 374–383.
Hamermesh, D.S., Frazis, H., Stewart, J., 2005. Data watch: the American time use survey. J. Econ. Perspect. 19 (1), 221–232.
Hammitt, J., et al., 2009. The effect of health and longevity on financial risk tolerance. Geneva Risk Ins. Rev. 34, 117–139.
Hamplova, D., Le Bourdais, C., 2009. One or two pot strategies? Income pooling in married and unmarried households in comparative perspective. J. Comp. Fam. Stud. 40, 355–385.
Hansen, K., Joshi, H., Verropoulou, G., 2006. Childcare and mothers' employment: approaching the millennium. Natl. Inst. Econ. Rev. 195, 84–102.
Harbury, C.D., Hitchens, D.M., 1977. Women, wealth and inheritance. Econ. J. 87 (345), 124–131.
Harkness, S., 2013. Women's employment and household income inequality. In: Gornick, J., Jantii, M. (Eds.), Income Inequality: Economic Disparities and the Middle Class in Affluent Countries. Stanford University Press, Stanford, CA, pp. 207–233.
Harkness, S., Waldfogel, J., 2003. The family gap in pay: evidence from seven industrialized countries. Res. Labor Econ. 22, 369–413.
Heckman, J., 1974. Shadow prices, market wages, and labor supply. Econometrica 42 (4), 679–694.
Heckman, J.J., 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. Ann. Econ. Soc. Meas. 5 (4), 475–492.
Heckman, J., 1979. Sample selection bias as a specification error. Econometrica 47 (1), 153–161.
Hegewisch, A., Liepmann, H., Hayes, J., Hartmann, H., 2010. Separate and not equal? Gender segregation in the labor market and the gender wage gap. Institute for Women's Policy Research Briefing Paper No. 377, Washington, DC.
Heikel, N., Liefbroer, A.C., Poortman, A.-R., 2010. Income pooling strategies among cohabiting and married couples: a comparative perspective. In: European Population Conference, Session 15, Vienna, 1–4, September.
Heimdal, K.R., Houseknecht, S.K., 2003. Cohabiting and married couples' income organization: approaches in Sweden and the United States. J. Marriage Fam. 65 (3), 525–538.
Hellerstein, J., Neumark, D., McInerney, M., 2007. Changes in workplace segregation in the United States between 1990 and 2000: evidence from matched employer–employee data. NBER Working Paper No. 13080.
Herr, J.L., Wolfram, C., 2009. Work environment and 'Opt-Out' rates at motherhood across high-education career paths. NBER Working Paper No. w14717.
Hersch, J., 1991. Male–female differences in hourly wages: the role of human capital, working conditions, and housework. Ind. Labor Relat. Rev. 44 (4), 746–759.
Hersch, J., Stratton, L.S., 2002. Housework and wages. J. Hum. Resour. 37 (1), 217–229.
Heyman, F., Svaleryd, H., Vlachos, J., 2013. Competition, takeovers, and gender discrimination. Ind. Labor Relat. Rev. 66, 409–511.
Hill, M.S., 1979a. The wage effects of marital status and children. J. Hum. Resour. 14 (4), 579–594.
Hill, T.P., 1979b. Do-it-yourself and GDP. Rev. Income Wealth 25 (1), 31–39.

Hiller, D.V., 1984. Power, dependence and division of family work. Sex Roles 10, 1003–1019.
Himmelweit, S., Santos, C., Sevilla, A., Sofer, C., 2013. Sharing of resources within the family and the economics of household decision making. J. Marriage Fam. 75 (3), 625–639.
Hipple, S.F., 2010. Self-employment in the United States. Mon. Labor Rev. 133 (9), 17–32
Hirsch, B., Konietzko, T., 2013. The effect of housework on wages in Germany: no impact at all. J. Labor Market Res. 46 (2), 103–118.
Hirsch, B., Schank, T., Schnabel, C., 2010. Differences in labor supply to monopsonistic firms and the gender pay gap: an empirical analysis using linked employer–employee data from Germany. J. Labor Econ. 28 (2), 291–330.
Hochschild, A., Machung, A., 1989. The Second Shift: Working Parents and the Revolution at Home. Viking, New York, NY.
Holzer, H.J., Neumark, D., 2000. What does affirmative action do? Ind. Labor Relat. Rev. 53 (2), 240–271.
Hook, J.L., 2006. Care in context: men's unpaid work in 20 countries, 1965–2003. Am. Sociol. Rev. 71 (4), 639–660.
Hook, J.L., 2010. Gender inequality in the welfare state: sex segregation in housework, 1965–2003. Am. J. Sociol. 115, 1480–1523.
Hotchkiss, J.L., Pitts, M.M., 2007. The role of labor market intermittency in explaining gender wage differentials. Am. Econ. Rev. 97 (2), 417–421.
Huerta, M.D.C., Adema, W., Baxter, J., Han, W.J., Lausten, M., Lee, R., Waldfogel, J., 2013. Fathers' leave, fathers' involvement and child development: are they related? Evidence from four OECD countries. p. 2013OECD Social Employment and Migration Working Papers No. 140.
Hundley, G., 2000. Male/female earnings differences in self-employment: the effects of marriage, children, and the household division of labor. Ind. Labor Relat. Rev. 54 (1), 95–104.
Hundley, G., 2001. Domestic division of labor and self/organizationally employed differences in job attitudes and earnings. J. Fam. Econ. Iss. 22 (2), 121–139.
Hunt, J., 2002. The transition in East Germany: when is a ten-point fall in the gender wage gap bad news? J. Labor Econ. 20 (1), 148–169.
Ironmonger, D., 1996. Counting outputs, capital inputs and caring labor: estimating gross household product. Fem. Econ. 2 (3), 37–64.
Ironmonger, D., 2000. Household production and the household economy. Research Paper, Department of Economics, University of Melbourne.
Ishida, J., 2003. The role of intrahousehold bargaining in gender discrimination. Ration. Soc. 15 (3), 361–380.
Jaumotte, F., 2003. Female labour force participation: past trends and main determinants in OECD countries. OECD Economics Department Working Papers No. 376.
Jefferson, T., 2005. Women and retirement incomes in Australia: a review. Econ. Rec. 81 (254), 273–291.
Jefferson, T., 2009. Women and retirement pensions: a research review. Fem. Econ. 15 (4), 115–145.
Jenkins, S.P., 1991. Poverty measurement and the within-household distribution: agenda for action. J. Soc. Policy 20 (4), 457–483.
Joshi, H., Macran, S., Dex, S., 1996. Employment after childbearing and women's subsequent labour force participation: evidence from the British 1958 birth cohort. J. Popul. Econ. 9 (3), 325–348.
Joshi, H., Paci, P., Waldfogel, J., 1999. The wages of motherhood: better or worse? Camb. J. Econ. 23 (5), 543–564.
Juhn, C., Murphy, K.M., 1997. Wage inequality and family labor supply. J. Labor Econ. 15 (1), 72–97 (part 1).
Juhn, C., Murphy, K.M., Pierce, B., 1991. Accounting for the slowdown in black–white wage convergence. In: Kosters, M.H. (Ed.), Workers and Their Wages. AEI Press, Washington, DC, pp. 107–143.
Juhn, C., Murphy, K.M., Pierce, B., 1993. Wage inequality and the rise in returns to skill. J. Polit. Econ. 101 (3), 410–442.
Jurajda, Š., Harmgart, H., 2007. When do female occupations pay more? J. Comp. Econ. 35 (1), 170–187.
Juster, E.T., Kuester, K.A., 1991. Differences in the measurement of wealth, wealth inequality and wealth composition obtained from alternative US wealth surveys. Rev. Income Wealth 37, 33–62.

Juster, F.T., Stafford, F.P., 1991. The allocation of time: empirical findings, behavioral models, and problems of measurement. J. Econ. Lit. 29 (2), 471–522.
Juster, F.T., Courant, P.N., Dow, G.K., 1981. A theoretical framework for the measurement of well-being. Rev. Income Wealth 27 (1), 1–31.
Kabátek, J., van Soest, A., Stancanelli, E., 2014. Income taxation, labour supply and housework: a discrete choice model for French couples. Labour Econ. 27, 30–43.
Kahn, J.R., García-Manglano, J., Bianchi, S.M., 2014. The motherhood penalty at midlife: long term effects of children on women's careers. J. Marriage Fam. 76 (1), 56–72.
Kahneman, D., Krueger, A.B., Schkade, D.A., Schwarz, N., Stone, A.A., 2004. Toward national well-being accounts. Am. Econ. Rev. 94 (2), 429–434.
Kalenkoski, C.M., Ribar, D.C., Stratton, L.S., 2009. The influence of wages on parents' allocations of time to child care and market work in the United Kingdom. J. Popul. Econ. 22 (2), 399–419.
Kalugina, E., Radtchenko, N., Sofer, C., 2009. How do spouses share their full income? Identification of the sharing rule using self-reported income. Rev. Income Wealth 55 (2), 360–391.
Kamo, Y., 1988. Determinants of household division of labor resources, power, and ideology. J. Fam. Issues 9 (2), 177–200.
Kamo, Y., 2000. He said, she said: assessing discrepancies in husbands' and wives' reports on the division of household labor. Soc. Sci. Res. 29, 459–476.
Kan, M.Y., 2008. Measuring housework participation: the gap between 'stylised' questionnaire estimates and diary-based estimates. Soc. Indic. Res. 86 (3), 381–400.
Kan, M.Y., Laurie, H., 2013. Changing patterns in the allocation of savings, investments and debts within couple relationships. Sociol. Rev. 62 (2), 335–358.
Kanbur, R., Haddad, L., 1994. Are better off households more unequal or less unequal? Oxf. Econ. Pap. 46 (3), 445–458.
Kassenboehmer, S.C., Sinning, M.G., 2014. Distributional changes in the gender wage gap. Ind. Labor Relat. Rev. 67 (2), 335–361.
Katz, E., 1997. The intra-household economics of voice and exit. Fem. Econ. 3 (3), 25–46.
Kenney, C., 2006. The power of the purse: allocative systems and inequality in couple households. Gend. Soc. 20 (3), 354–381.
Kihlstrom, R.E., Laffont, J.J., 1979. A general equilibrium entrepreneurial theory of firm formation based on risk aversion. J. Polit. Econ. 87 (4), 719–748.
Killewald, A., 2011. Opting out and buying out: wives' earnings and housework time. J. Marriage Fam. 73 (2), 459–471.
Killewald, A., 2013. A reconsideration of the fatherhood premium marriage, coresidence, biology, and fathers' wages. Am. Sociol. Rev. 78 (1), 96–116.
Killewald, A., Gough, M., 2010. Money isn't everything: wives' earnings and housework time. Soc. Sci. Res. 39 (6), 987–1003.
Killingsworth, M.R., 1990. The Economics of Comparable Worth. W.E. Upjohn Institute for Employment Research, Kalamazoo.
Killingsworth, M.R., Heckman, J.J., 1986. Female labor supply: a survey. In: Ashenfelter, O., Layard, R. (Eds.), Handbook of Labor Economics, vol. 1. Elsevier Science, Amsterdam, pp. 103–204.
Kim, J.W., Choi, Y.J., 2013. Feminisation of poverty in 12 welfare states: consolidating cross-regime variations? Int. J. Soc. Welfare 22 (4), 347–359.
Kim, M.K., Polachek, S.W., 1994. Panel estimates of male–female earnings functions. J. Hum. Resour. 29 (2), 406–428.
Kirzner, I.M., 1979. Perception, Opportunity, and Profit: Studies in the Theory of Entrepreneurship. University of Chicago Press, Chicago, IL.
Klein, M.W., Moser, C., Urban, D.M., 2010. The contribution of trade to wage inequality: the role of skill, gender, and nationality. NBER Working Paper Series No. 15985.
Klerman, J.A., Leibowitz, A., 1999. Job continuity among new mothers. Demography 36 (2), 145–155.
Kleven, H.J., Kreiner, C.T., Saez, E., 2009. The optimal income taxation of couples. Econometrica 77 (2), 537–560.

Kluve, J., Tamm, M., 2013. Parental leave regulations, mothers' labor force attachment and fathers' childcare involvement: evidence from a natural experiment. J. Popul. Econ. 26, 983–1005.
Knight, F.H., 1921. Risk, Uncertainty and Profit. Houghton Mifflin, Boston, p. 381.
Koellinger, P., Minniti, M., Schade, C., 2013. Gender differences in entrepreneurial propensity. Oxf. Bull. Econ. Stat. 75 (2), 213–234.
Kollmeyer, C., 2013. Family structure, female employment, and national income inequality: a cross-national study of 16 Western countries. Eur. Sociol. Rev. 29 (4), 816–827.
Korenman, S., Neumark, D., 1991. Does marriage really make men more productive? J. Hum. Resour. 16 (2), 282–307.
Korenman, S., Neumark, D., 1992. Marriage, motherhood, and wages. J. Hum. Resour. 27 (2), 233–255.
Kotsadam, A., 2011. Does informal eldercare impede women's employment? The case of European welfare states. Fem. Econ. 17 (2), 121–144.
Kreyenfeld, M., Hank, K., 2000. Does the availability of child care influence the employment of mothers? Findings from western Germany. Popul. Res. Policy Rev. 19 (4), 317–337.
Krueger, A.B., 2007. Are we having more fun yet? Categorizing and evaluating changes in time allocation. Brook. Pap. Econ. Act. 2007 (2), 193–215.
Krueger, A.B., Kahneman, D., Schkade, D.A., Schwarz, N., Stone, A.A., 2009. National time accounting: the currency of life. In: Krueger, A.B. (Ed.), Measuring the Subjective Well-Being of Nations: National Accounts of Time Use and Well-Being. NBER, pp. 9–86. Available from: http://www.nber.org/chapters/c5053.
Kuhn, P., Villeval, M.C., 2013. Are women more attracted to cooperation than men? NBER Working Paper No. 19277.
Kurtulus, F.A., Tomaskovic-Devey, D., 2012. Do female top managers help women to advance? A panel study using EEO-1 records. Ann. Am. Acad. Pol. Soc. Sci. 639 (1), 173–197.
Lalive, R., Zweimüller, J., 2009. How does parental leave affect fertility and return to work? Evidence from two natural experiments. Q. J. Econ. 124 (3), 1363–1402.
Landefeld, J.S., McCulla, S.H., 2000. Accounting for nonmarket household production within a national accounts framework. Rev. Income Wealth 46 (3), 289–307.
Landefeld, J.S., Fraumeni, B.M., Vojtech, C.M., 2009. Accounting for nonmarket production: a prototype satellite account using the American time use survey. Rev. Income Wealth 55 (2), 205–225.
Langowitz, N., Minniti, M., 2007. The entrepreneurial propensity of women. Enterp. Theory Pract. 31 (3), 341–364.
Laporte, C., Schellenberg, G., 2011. The income management strategies of older couples in Canada. Can. Stud. Popul. 38 (3–4), 1–22.
Lavy, V., 2012. Gender differences in market competitiveness in a real workplace: evidence from performance-based pay tournaments among teachers. NBER Working Paper No. 14338.
Lazear, E.P., 2005. Entrepreneurship. J. Labor Econ. 23 (4), 649–680.
Lazear, E.P., Rosen, S., 1990. Male–female wage differentials in job ladders. J. Labor Econ. 8 (1), S106–S123.
Lechene, V., 1993. Une revue de la littérature sur les échelles d'équivalence. Econ. Prévision 110–111, 169–182.
Lechmann, D.S., Schnabel, C., 2012. Why is there a gender earnings gap in self-employment? A decomposition analysis with German data. J. Eur. Labor Stud. 1 (1), 1–25.
Lee, Y.S., Waite, L.J., 2005. Husbands' and wives' time spent on housework: a comparison of measures. J. Marriage Fam. 67 (2), 328–336.
Leibbrandt, A., List, J., 2012. Do women avoid salary negotiations? Evidence from a large scale natural field experiment. NBER Working Paper No. 18511.
Leonard, J., 1989. Women and affirmative action. J. Econ. Perspect. 3 (1), 61–75.
Leoni, T., Falk, M., 2010. Gender and field of study as determinants of self-employment. Small Bus. Econ. 34 (2), 167–185.
Lequien, L., 2012. The impact of parental duration on later wages. Ann. Econ. Stat. 107, 267–286.
Leung, D., 2006. The male/female earnings gap and female self-employment. J. Socio-Econ. 35 (5), 759–779.

Levitt, S.D., List, J.A., 2007. What do laboratory experiments measuring social preferences reveal about the real world? J. Econ. Perspect. 21 (2), 153–174.

Lewbel, A., Pendakur, K., 2008. Equivalence scales, second ed. The new palgrave dictionary of economics. In: Durlauf, S.N., Blume, L.E. (Eds.), The New Palgrave Dictionary of Economics. Palgrave Macmillan, Basingstoke, Hampshire New York.

Lilly, M.B., Laporte, A., Coyte, P.C., 2007. Labor market work and home care's unpaid caregivers: a systematic review of labor force participation rates, predictors of labor market withdrawal, and hours of work. Milbank Q. 85 (4), 641–690.

Lincoln, A.E., 2008. Gender, productivity, and the marital wage premium. J. Marriage Fam. 70 (3), 806–814.

Lise, J., Seitz, S., 2011. Consumption inequality and intra-household allocations. Rev. Econ. Stud. 78, 328–355.

Livermore, T., Rodgers, J., Siminski, P., 2011. The effect of motherhood on wages and wage growth: evidence for Australia. Econ. Rec. 87 (s1), 80–91.

Lombard, K.V., 2001. Female self-employment and demand for flexible, non-standard work schedules. Econ. Inq. 29 (2), 214–317.

Loughran, D.S., Zissimopoulos, J.M., 2009. Why wait? The effect of marriage and childbearing on the wages of men and women. J. Hum. Resour. 44 (2), 326–349.

Loury, L.D., 1997. Gender earnings gap among college-educated workers. Ind. Labor Relat. Rev. 50 (4), 580–593.

Love, D.A., 2010. The effects of marital status and children on savings and portfolio choice. Rev. Financ. Stud. 23 (1), 385–432.

Lucifora, C., Meurs, D., 2006. The public sector pay gap in France, Great Britain and Italy. Rev. Income Wealth 52 (1), 43–59.

Luckhaus, L., 2000. Equal treatment, social protection and income security for women. Int. Labor Rev. 139 (2), 149–178.

Ludwig-Mayerhofer, W., Gartner, H., Allmendinger, J., 2006. The allocation of money in couples: the end of inequality? Zeisch. Soziol. 35 (3), 212–226.

Lundberg, S., 2008. Gender and household decision making. In: Bettio, F., Veraschagina, A. (Eds.), Frontiers in the Economics of Gender. Routledge, London, pp. 116–134.

Lundberg, S., Pollak, R.A., 1993. Separate spheres bargaining and the marriage market. J. Polit. Econ. 100 (6), 988–1010.

Lundberg, S., Pollak, R.A., 2013. Cohabitation and the uneven retreat from marriage in the US, 1950–2010. In: Human Capital in History: The American Record, NBER Working Papers No19413, National Bureau of Economic Research.

Lundberg, S., Rose, E., 2000. Parenthood and the earnings of married men and women. Labor Econ. 7 (6), 689–710.

Lundberg, S., Pollack, R.A., Wales, T., 1997. Do husbands and wives pool their resources? Evidence from the UK child benefit. J. Hum. Resour. 32 (3), 463–480.

Lundin, D., Mörk, E., Öckert, B., 2008. How far can reduced childcare prices push female labour supply? Labour Econ. 15 (4), 647–659.

Lupton, J.P., Smith, J.P., 2003. Marriage, assets and savings. In: Grossbard-Shechtman, S.A. (Ed.), Marriage and the Economy: Theory and Evidence from Advanced Industrial Societies. Cambridge University Press, Cambridge, pp. 129–152.

Lyngstad, T.H., Noack, T., Tufte, P.A., 2011. Pooling of economic resources: a comparison of Norwegian married and cohabiting couples. Eur. Sociol. Rev. 27 (5), 624–635.

Maani, A.A., Cruickshank, A.A., 2010. What is the effect of housework on the market wage, and can it explain the gender wage gap? J. Econ. Surv. 24 (3), 402–427.

MacDonald, M., Phipps, S., Lethbridge, L., 2005. Taking its toll: the influence of paid and unpaid work on women's well-being. Fem. Econ. 11 (1), 63–94.

Machado, J.F., Mata, J., 2005. Counterfactual decomposition of changes in wage distributions using quantile regression. J. Appl. Econ. 20, 445–465.

Machin, S., Puhani, P.A., 2003. Subject of degree and the gender wage differential: evidence from the UK and Germany. Econ. Lett. 79 (3), 393–400.

Mackenzie, G.A.S., 2010. The Decline of the Traditional Pension. Cambridge University Press, United States of America, p. 275.
MacPherson, D.A., 1988. Self-employment and married women. Econ. Lett. 28 (3), 281–284.
MacPherson, D.A., Hirsch, B.T., 1995. Wages and gender composition: why do women's jobs pay less? J. Labor Econ. 13 (3), 426–471.
Maitre, B., Nolan, B., Whelan, C.T., 2012. Low pay, in-work poverty and economic vulnerability: a comparative analysis using EU-SILC. Manch. Sch. 80 (1), 99–116.
Malkiel, B.G., Malkiel, J.A., 1973. Male–female pay differentials in professional employment. Am. Econ. Rev. 63 (4), 693–705.
Manchester, C.F., Leslie, L.M., Kramer, A., 2010. Stop the clock policies and career success in academia. Am. Econ. Rev. 100 (2), 219–223.
Mandel, H., Semyonov, M., 2006. A welfare state paradox: state interventions and women's employment opportunities in 22 countries. Am. J. Sociol. 111 (6), 1910–1949.
Manning, A., 2003. Monopsony in Motion. Princeton University Press, Princeton NJ, p. 416.
Manning, A., Petrongolo, B., 2008. The part-time pay penalty for women in Britain. Econ. J. 118 (526), F28–F51.
Manning, A., Saidi, F., 2010. Understanding the gender pay gap: what's competition got to do with it? Ind. Labor Relat. Rev. 63 (4), 681–698.
Manning, A., Swaffield, J., 2008. The gender gap in early-career wage growth. Econ. J. 118 (530), 983–1024.
Manser, M., Brown, M., 1980. Marriage and household decision-making: a bargaining approach. Int. Econ. Rev. 21 (1), 31–44.
Marshall, M.I., Flaig, A., 2014. Marriage, children, and self-employment earnings: an analysis of self-employed women in the US. J. Fam. Econ. Iss. 35 (3), 313–322.
Matsa, D., Miller, A., 2011. Chipping away at the glass ceiling: gender spillovers in corporate leadership. Am. Econ. Rev. 101 (3), 635–639.
Matsa, D.A., Miller, A.R., 2013. A female style in corporate leadership? Evidence from quotas. Am. Econ. J. 5 (3), 136–169.
Mattingly, M.J., Blanchi, S.M., 2003. Gender differences in the quantity and quality of free time: the US experience. Soc. Forces 81 (3), 999–1030.
Maume, D.J., 2006. Gender differences in restricting work efforts because of family responsibilities. J. Marriage Fam. 68 (4), 859–869.
McElroy, M.B., Horney, M.J., 1981. Nash bargained household decisions: toward a generalization of the theory of demand. Int. Econ. Rev. 22 (2), 333–349.
McLennan, M.C., 2000. Does household labour impact market wages? Appl. Econ. 32 (12), 1541–1557.
McRae, S., 2003. Constraints and choice in mother's employment careers: a consideration of Hakim's preference theory. Br. J. Sociol. 54 (3), 317–388.
Mehay, S.L., Bowman, W.R., 2005. Marital status and productivity: evidence from personnel data. South. Econ. J. 72 (1), 63–77.
Melly, B., 2005. Public–private sector wage differentials in Germany: evidence from quantile regression. Empir. Econ. 30 (2), 505–520.
Mencarini, L., Sironi, M., 2012. Happiness, housework and gender inequality in Europe. Eur. Sociol. Rev. 28 (2), 203–219.
Meng, X., Meurs, D., 2004. The gender earnings gap: effects of institutions and firms—a comparative study of French and Australian private firms. Oxf. Econ. Pap. 56 (2), 189–208.
Meulders, D., O'Dorchai, S., 2010. Revisiting poverty measures towards individualisation. ULB-Dulbea Working Paper No. 10-03.
Meurs, D., Ponthieux, S., 2008. Public and private employment and the gender wage gap in 8 European countries. In: ETUI, Privatisation and Marketisation of Services, pp. 261–283. Brussels, ETUI-REHS.
Miller, A.R., 2011. The effects of motherhood timing on career path. J. Popul. Econ. 24 (3), 1071–1100.
Miller, T., del Carmen Triana, M., 2009. Demographic diversity in the boardroom: mediators of the board diversity–firm performance relationship. J. Manag. Stud. 46 (5), 755–786.
Mincer, J., 1962. Labor force participation of married women: a study of labor supply. In: NBER Volume, (Ed.), Aspects of Labor, Economics, pp. 63–106. http://www.nber.org/chapters/c0603.

Mincer, J., Ofek, H., 1982. Interrupted work careers: depreciation and restoration of human capital. J. Hum. Resour. 17 (1), 3–24.
Mincer, J., Polachek, S., 1974. Family investments in human capital: earnings of women. J. Polit. Econ. 82 (2), S76–S108.
Minniti, M., Nardone, C., 2007. Being in someone else's shoes: the role of gender in nascent entrepreneurship. Small Bus. Econ. 28 (2–3), 223–238.
Miranda, V., 2011. Cooking, caring and volunteering: unpaid work around the world. OECD Social, Employment and Migration Working Papers 116. OECD Publishing.
Misra, J., Budig, M., Boeckmann, I., 2011. Work-family policies and the effects of children on women's employment hours and wages. Commun. Work Fam. 14 (2), 139–157.
Molina, J.A., Montuenga, V.M., 2009. The motherhood wage penalty in Spain. J. Fam. Econ. Iss. 30 (3), 237–251.
Moss, M., 1973. Introduction. In: Moss, M. (Ed.), The Measurement of Economic and Social Performance. NBER, pp. 1–22.
Moss-Racusin, C., Dovidio, J., Brescoli, V., Graham, M., Handelsma, J., 2012. Science faculty's subtle gender biases favor male students. Proc. Natl. Acad. Sci. U.S.A. 109 (41), 16474–16479.
Nakosteen, R.A., Zimmer, M.A., 1997. Men, money and marriage: are high earners more prone than low earners to marry? Soc. Sci. Q. 78 (1), 66–82.
Neelakantan, U., 2010. Estimation and impact of gender differences in risk tolerance. Econ. Inq. 48 (1), 228–233.
Neelakantan, U., Lyons, A., Nelson, C., 2009. Household bargaining and portfolio choice. Working Paper, University of Illinois at Urbana-Champaign. www.aeaweb.org/assa/2009/retrieve.php.
Nelson, J.A., 1994. I, thou and them: capabilities, altruism and norms in the economics of marriage. Am. Econ. Rev. 84 (2), 126–131.
Nepomnyaschy, L., Waldfogel, J., 2007. Paternity leave and fathers' involvement with their young children: evidence from the American Ecls-B. Commun. Work Fam. 10 (4), 427–453.
Neuman, S., Oaxaca, R.L., 1998. Estimating labour market discrimination with selectivity corrected wage equations: methodological considerations and an illustration from Israel. Centre for Economic Policy Research, Working Paper No. 1915.
Neumark, D., 1988. Employers' discriminatory behavior and the estimation of wage discrimination. J. Hum. Resour. 23 (3), 279–295.
Neumark, D.M., 1996. Sex discrimination in restaurant hiring: an audit study. Q. J. Econ. 111, 915–941.
Neumark, D., Korenman, S., 1992. Sources of bias in women's wage equations: results using sibling data. NBER Working Paper No. w4019.
Ngai, L.R., Petrongolo, B., 2013. Gender gaps and the rise of the service economy. CEPR Discussion Paper 1204.
Niederle, M., Vesterlund, L., 2007. Do women shy away from competition? Do men compete too much? Q. J. Econ. 122 (3), 1067–1101.
Niederle, M., Vesterlund, L., 2011. Gender and competition. Annu. Rev. Econ. 3 (1), 601–630.
Niederle, M., Segal, C., Vesterlund, L., 2013. How costly is diversity? Affirmative action in light of gender differences in competitiveness. Manag. Sci. 59 (1), 1–16.
Noonan, M.C., 2001. The impact of domestic work on men's and women's wages. J. Marriage Fam. 63 (4), 1134–1145.
Ñopo, H., 2008a. Matching as a tool to decompose wage gaps. Rev. Econ. Stat. 90 (2), 290–299.
Ñopo, H., 2008b. An extension of the Blinder–Oaxaca decomposition to a continuum of comparison groups. Econ. Lett. 100 (2), 292–296.
Nordhaus, W.D., Tobin, J., 1973. Is growth obsolete? In: Moss, M. (Ed.), The Measurement of Economic and Social Performance. NBER, pp. 509–564. http://www.nber.org/chapters/c3621.
Oaxaca, R.L., 1973. Male–female wage differentials in urban labor markets. Int. Econ. Rev. 14 (3), 693–709.
Oaxaca, R.L., Ransom, M.R., 1994. On discrimination and the decomposition of wage differentials. J. Econ. 61 (1), 5–21.
OECD, 2007. Babies and Bosses: Reconciling Work and Family Life. A Synthesis of Findings for OECD Countries. OECD, Paris.

OECD, 2011a. Society at a Glance 2011: OECD Social Indicators. OECD Publishing, Paris.
OECD, 2011b. Doing Better for Families. OECD, Paris.
OECD, 2012. Closing the Gap—Act Now. OECD Publishing. Data available from, http://www.oecd.org/gender/data/balancingpaidworkunpaidworkandleisure.htm.
OECD, 2013. The OECD Framework for Statistics on the Distribution of Household Income, Consumption and Wealth. OECD Publishing, Paris.
OECD/The European Commission, 2013. The Missing Entrepreneurs: Policies for Inclusive Entrepreneurship in Europe. OECD Publishing, Paris.
Offer, S., Schneider, B., 2011. Revisiting the gender gap in time-use patterns multitasking and well-being among mothers and fathers in dual-earner families. Am. Sociol. Rev. 76 (6), 809–833.
Olivetti, C., 2006. Changes in women's hours of market work: the role of returns to experience. Rev. Econ. Dyn. 9 (4), 557–587.
Olivetti, C., Petrongolo, B., 2008. Unequal pay or unequal employment? A cross-country analysis of gender gaps. J. Labor Econ. 26 (4), 621–654.
Olsberg, D., 2006. Major changes for super, but women are still Ms. . .ing out. Aust. Account. Rev. 16 (40), 47–51.
O'Neil, J., Polacheck, S., 1993. Why the gender gap in wages narrowed in the 1980's. J. Labor Econ. 11 (1), 205–208.
Oostveen, A., Biletta, I., Parent-Thirion, A., Vermeylen, G., 2013. Self-employed or not self-employed? Working conditions of economically dependent workers. Background Paper, Eurofound.
Orenstein, M.A., 2013. Pension privatization: evolution of a paradigm. Governance 26 (2), 259–281.
Ozawa, M.N., Lee, Y., 2006. The net worth of female-headed households: a comparison to other types of households. Fam. Relat. 55 (1), 132–145.
Pahl, J., 1983. The allocation of money and the structuring of inequality within marriage. Sociol. Rev. 31, 235–262.
Pahl, J., 1989. Money and Marriage. Macmillan Education, Basingstoke, UK.
Pahl, J., 1995. His money, her money: recent research on financial organization in marriage. J. Econ. Psychol. 16, 361–376.
Pande, R., Ford, D., 2012. Gender quotas and female leadership. World Development Report, World Bank.
Peña-Casas, R., Ghailani, D., 2011. Towards individualizing gender in-work poverty risks. In: Fraser, N. et al., (Ed.), Working Poverty in Europe: A Comparative Approach. Palgrave Macmillan, London.
Pekkarinen, T., Vartiainen, J., 2006. Gender differences in promotion on a job ladder: evidence from Finnish metalworkers. Ind. Labor Relat. Rev. 59 (2), 285–301.
Perry, S.C., 2002. A comparison of failed and non-failed small business in the United States: do men and women use different planning and decision making strategies? J. Dev. Entrep. 7 (4), 415.
Pestieau, P., 2003. The role of gift and estate transfers in the United States and in Europe. In: Munnell, A., Sunden, A. (Eds.), Death and Dollars. The Brookings Institution Press, Washington, DC, pp. 64–85.
Petersen, T., Morgan, L.A., 1995. Separate and unequal: occupation-establishment sex segregation and the gender wage gap. Am. J. Sociol. 101 (2), 329–365.
Petersen, T., Penner, A.M., Hogsnes, G., 2011. Male marital wage premium: sorting vs. differential pay. Ind. Labor Relat. Rev. 64 (2), 283–304.
Petit, P., 2007. The effects of age and family constraints on gender hiring discrimination: a field experiment in the French financial sector. Labour Econ. 14 (3), 371–391.
Pfeifer, C., 2011. Risk aversion and sorting into public sector employment. Ger. Econ. Rev. 12 (1), 85–99.
Phelps, E.S., 1972. The statistical theory of racism and sexism. Am. Econ. Rev. 62 (4), 659–661.
Phipps, S.A., Burton, P.S., 1995. Sharing within families: implications for the measurement of poverty among individuals in Canada. Can. J. Econ. 28 (1), 177–204.
Phipps, S.A., Burton, P.S., 1998. What's mine is yours? The influence of male and female incomes on patterns of household expenditure. Economica 65 (260), 599–613.
Phipps, S., Woolley, F., 2008. Control over money and the savings decisions of Canadian households. J. Socio-Econ. 37 (2), 592–611.
Phipps, S., Burton, P., Osberg, L., 2001a. Time as a source of inequality within marriage: are husbands more satisfied with time for themselves than wives? Fem. Econ. 7 (2), 1–21.

Phipps, S., Burton, P., Lethbridge, L., 2001b. In and out of the labour market: long-term consequences of child-related interruptions of women's paid work. Can. J. Econ. 34 (2), 411–429.
Piggott, J., Whalley, J., 1996. The tax unit and household production. J. Polit. Econ. 104 (2), 398–418.
Piketty, T., 2013. Le capital au XXIe siècle. Editions du Seuil, Paris Also [2014]. Capital in the Twenty-First Century. Harvard University Press.
Plug, E.J., Van Praag, B.M., 1998. Similarity in response behavior between household members: an application to income evaluation. J. Econ. Psychol. 19, 497–513.
Poissonnier, A., Roy, D., 2013. Households satellite account for France in 2010. Methodological issues on the assessment of domestic production. INSEE Working Paper No. g2013–14.
Polachek, S.W., 1981. Occupational self-selection: a human capital approach to sex differences in occupational structure. Rev. Econ. Stat. 63 (1), 60–69.
Polachek, S.W., 2004. How the human capital model explains why the gender wage gap narrowed. IZA Discussion Paper Series No. 1102.
Pollak, R.A., 1985. A transaction cost approach to families and households. J. Econ. Lit. 23 (2), 581–608.
Pollak, R.A., 2003. Gary Becker's contributions to family and household economics. Rev. Econ. Househ. 1 (1), 111–141.
Pollak, R.A., 2005. Bargaining power in marriage: earnings, wage rates and household production. NBER Working Paper No. 11239.
Pollak, R.A., 2013. Allocating household time: when does efficiency imply specialization? NBER Working Paper No. 19178.
Pollak, R.A., Wachter, M.L., 1975. The relevance of the household production function and its implications for the allocation of time. J. Polit. Econ. 83 (2), 255–278.
Ponthieux, S., 2010. Assessing and analysing in-work poverty risk. In: Atkinson, A.B., Marlier, E. (Eds.), Income and Living Conditions in Europe. EU Publications Office, Luxembourg.
Ponthieux, S., 2013. Income pooling and equal sharing within the household—what can we learn from the 2010 EU-Silc module. Eurostat Working Papers and Methodological Series, Luxembourg.
Pronzato, C.D., 2009. Return to work after childbirth: does parental leave matter in Europe? Rev. Econ. Househ. 7 (4), 341–360.
Ramsey, F.P., 1927. A contribution to the theory of taxation. Econ. J. 47–61.
Rangel, M.A., 2006. Alimony rights and intrahousehold allocation of resources: evidence from Brazil. Econ. J. 116 (513), 627–658.
Ransom, M., Oaxaca, R.L., 2005. Intrafirm mobility and sex differences in pay. Ind. Labor Relat. Rev. 28 (2), 219–237.
Ransom, M., Oaxaca, R., 2010. New market power and sex differences in pay. J. Labor Econ. 28 (2), 267–290.
Regan, T.L., Oaxaca, R.L., 2009. Work experience as a source of specification error in earnings models: implications for gender wage decompositions. J. Popul. Econ. 22 (2), 463–499.
Reid, M.G., 1934. Economics of Household Production. Wiley, New York.
Reimers, C.W., 1983. Labor market discrimination against Hispanic and Black men. Rev. Econ. Stat. 65 (4), 570–579.
Régnier-Loilier, A., 2009. Does the birth of a child change the division of household tasks between partners? Population and Societies No. 461.
Riach, P.A., Rich, J., 2002. Field experiments of discrimination in the market place. Econ. J. 112 (483), 480–518.
Rosen, H.S., 1977. Is it time to abandon joint filing? Natl. Tax J. 30, 423–428.
Ruel, E., Hauser, R.M., 2013. Explaining the gender wealth gap. Demography 50 (4), 1155–1176.
Ruhm, C.J., 1998. The economic consequences of parental leave mandates: lessons from Europe. Q. J. Econ. 113 (1), 287–317.
Ruiz-Arranz, M., Davis, B., Handa, S., Stampini, M., Winters, P., 2006. Program conditionality and food security: the impact of PROGRESA and PROCAMPO transfers in rural Mexico. Rev. Econ. 7 (2), 249–278.

Ruppanner, L., 2010. Cross-national reports of housework: an investigation of the gender egalitarianism measure. Soc. Sci. Res. 39, 963–975.
Sabatelli, R.M., Shehan, C.L., 1993. Exchange and resource theories. In: Boss, P.G. et al., (Ed.), Sourcebook of Family Theories and Methods: A Contextual Approach. Plenum Press, New York, NY, pp. 385–411.
Safilios-Rothschild, C., 1970. The study of family power structure: a review 1960–1969. J. Marriage Fam. 32 (4), 539–552.
Safilios-Rothschild, C., 1976. A macro- and micro-examination of family power and love: an exchange model. J. Marriage Fam. 38 (2), 355–362.
Saint-Paul, G., 2008. Against gender-based taxation. CEPR Document Paper 6582.
Salverda, W., Checchi, D., 2014. Labour-market institutions and the dispersion of wage earnings. In: Atkinson, Bourguignon (Eds.), Handbook of Income Distribution, This volume (Chapter 18).
Samuelson, P.-A., 1956. Social indifference curves. Q. J. Econ. 70 (1), 1–22.
Sanders, C.K., Porterfield, S.L., 2010. The ownership society and women: exploring female householders' ability to accumulate assets. J. Fam. Econ. Iss. 31 (1), 90–106.
Saparito, P., Elam, A., Brush, C., 2013. Bank–firm relationships: do perceptions vary by gender? Enterp. Theory Pract. 37 (July), 837–858.
Saridakis, G., Marlow, S., Storey, D.J., 2013. Do different factors explain male and female self-employment rates? J. Bus. Ventur. 29 (3), 345–362.
Sayer, L.C., 2005. Gender, time and inequality: trends in women's and men's paid work, unpaid work and free time. Soc. Forces 84 (1), 285–303.
Sayer, L.C., 2010. Trends in housework. In: Treas, J., Drobnic, S. (Eds.), Dividing the Domestic—Men, Women, and Household Work in Cross-National Perspective. Stanford University Press, Stanford, CA, pp. 19–40.
Sayer, L.C., Bianchi, S.M., Robinson, J.P., 2004. Are parents investing less in children? Trends in mothers' and fathers' time with children. Am. J. Sociol. 110 (1), 1–43.
Sayer, L.C., England, P., Bittman, M., Bianchi, S.M., 2009. How long is the second (plus first) shift? Gender differences in paid, unpaid, and total work time in Australia and the United States. J. Comp. Fam. Stud. 40 (4), 523–545.
Schmidt, L., Sevak, P., 2006. Gender, marriage, and asset accumulation in the United States. Fem. Econ. 12 (1–2), 139–166.
Schneeweis, N., Zweimüller, M., 2012. Girls, girls, girls: gender composition and female school choice. Econ. Educ. Rev. 31 (4), 482–500.
Schneider, D., 2011. Market earnings and household work: new tests of gender performance theory. J. Marriage Fam. 73 (4), 845–860.
Sedo, S.A., Kossoudji, S.A., 2004. Rooms of one's own: gender, race and home ownership as wealth accumulation in the United States. IZA DP No. 1397.
Sena, V., Scott, J., Roper, S., 2012. Gender, borrowing patterns and self-employment: some evidence for England. Small Bus. Econ. 38 (4), 467–480.
Sevilla-Sanz, A., Gimenez-Nadal, J.I., Fernandez, C., 2010. Gender roles and the division of unpaid work in Spanish households. Fem. Econ. 16 (4), 137–184.
Shelton, B.A., 1992. Women, Men and Time: Gender Differences in Paid Work, Housework and Leisure. Greenwood Press, Westport, Conn.
Shilton, E., 2012. Insuring inequality: sex-based mortality tables and women's retirement income. Queen Law J. 37 (2), 383–731.
Shorrocks, A.F., 1982. The portfolio composition of asset holdings in the United Kingdom. Econ. J. 92 (366), 268–284.
Shrader, C.B., Blackburn, V.L., Iles, P., 1997. Women in management and firm financial performance: an exploratory study. J. Manage. Issues 9 (3), 355–372.
Sierminska, E., Smeeding, T., 2005. Measurement Issues: Equivalence Scales, Accounting Framework and Reference Unit. Luxembourg Income Study, Luxembourg.
Sierminska, E.M., Frick, J.R., Grabka, M.M., 2010. Examining the gender wealth gap. Oxf. Econ. Pap. 62 (4), 669–690.

Sigle-Rushton, W., Waldfogel, J., 2007. Motherhood and women's earnings in Anglo-American, Continental European, and Nordic countries. Fem. Econ. 13 (2), 55–91.

Simonsen, M., Skipper, L., 2006. The costs of motherhood: an analysis using matching estimators. J. Appl. Econ. 21 (7), 919–934.

Simonsen, M., Skipper, L., 2012. The family gap in wages: what wombmates reveal. Labour Econ. 19 (1), 102–112.

Smeeding, T.M., Weinberg, D.H., 2001. Toward a uniform definition of household income. Rev. Income Wealth 47 (1), 1–24.

Smith, J.D., 1974. The concentration of personal wealth in America, 1969. Rev. Income Wealth 20 (2), 143–180.

Smith, N., Smith, V., Verner, M., 2005. Do women in top management affect firm performance? A panel study of 2500 Danish firms. IZA DP No. 1708, 2005.

Smith, N., Smith, V., Verner, M., 2006. Do women in top management affect firm performance? A panel study of 2,500 Danish firms. Int. J. Product. Perform. Manag. 55 (7), 569–593.

Smith, N., Smith, V., Verne, M., 2011. The gender pay gap in top corporate jobs in Denmark: glass ceilings, sticky floors or both? Int. J. Manpow. 32 (2), 156–177.

Smith, N., Smith, V., Verner, M., 2013. Why are so few females promoted into CEO and Vice President Positions: Danish Empirical evidence, 1997–2007. Ind. Labor Relat. Rev. 66 (2), 380–408.

Sorensen, A., MacLanahan, S., 1987. Married women's economic dependency, 1940–1980. Am. J. Sociol. 93 (3), 659–687.

South, S.J., Spitze, G., 1994. Housework in marital and non-marital households. Am. Sociol. Rev. 59, 327–347.

Spiess, C.K., Schneider, A.U., 2003. Interactions between care-giving and paid work hours among European midlife women, 1994 to 1996. Ageing Soc. 23 (1), 41–68. Westport, CT: Greenwood.

Stevenson, B., Wolfers, J., 2009. The paradox of declining female happiness. Am. Econ. J. 1 (2), 190–225.

Stiglitz, J., Sen, A., Fitoussi, J.-P., 2009. Report by the Commission on the Measurement of Economic Performance and Social Progress. www.stiglitz-sen-fitoussi.fr.

Sullivan, O., 2011. An end to gender display through the performance of housework? A review and reassessment of the quantitative literature using insights from the qualitative literature. J. Fam. Theory Rev. 3 (1), 1–13.

Sullivan, O., Gershuny, J., 2001. Cross-national changes in time-use: some sociological (hi)stories re-examined. Br. J. Sociol. 52 (2), 331–347.

Sundström, M., Duvander, A.Z.E., 2002. Gender division of childcare and the sharing of parental leave among new parents in Sweden. Eur. Sociol. Rev. 18 (4), 433–447.

Sutherland, H., 1997. Women, men and the redistribution of income. Fisc. Stud. 18 (1), 1–22.

Sutter, M., Kocher, M.G., Glatzle-Rutzler, D., Trautmann, S.T., 2013. Impatience and uncertainty: experimental decisions predict adolescents' field behavior. Am. Econ. Rev. 103 (1), 510–531.

Tavora, I., 2012. The southern European social model: familialism and the high rates of female employment in Portugal. J. Eur. Soc. Pol. 22 (1), 63–76.

Thomas, D., 1990. Intra-household resource allocation: an inferential approach. J. Hum. Resour. 24 (4), 635–664.

Thévenon, O., 2011. Family policies in OECD countries: a comparative analysis. Popul. Dev. Rev. 37 (1), 57–87.

Thévenon, O., 2013. Drivers of female labour force participation in the OECD. http://dx.doi.org/10.1787/5k46cvrgnms6-en, OECD Social, Employment and Migration Working Papers No. 145, OECD Publishing.

Thévenon, O., Solaz, A., 2013. Labour market effects of parental leave policies in OECD countries. Directorate for Employment, Labour and Social Affairs, OECD.

Tichenor, V.J., 1999. Status and income as gendered resources. The case of marital power. J. Marriage Fam. 61 (3), 638–650.

Tison, A., Davin, B., Ventelou, B., Paraponaris, A., 2012. Influence of diseases on risk aversion through time. Working Paper, Aix Marseille School of Economics.

Treas, J., 1993. Money in the bank: transaction costs and the economic organization of marriage. Am. Sociol. Rev. 58, 723–734.
Treas, J., Lui, J., 2013. Studying housework across nations. J. Fam. Theory Rev. 5 (2), 135–149.
Treas, J., Tai, T.O., 2012. How couples manage the household: work and power in cross-national perspective. J. Fam. Issues 33 (8), 1088–1116.
Tymula, A., Belmaker, L.A.R., Ruderman, L., Glimcher, P.W., Levy, I., 2013. Like cognitive function, decision making across the life span shows profound age-related changes. Proc. Natl. Acad. Sci. U.S.A. 110 (42), 17143–17148.
Ulker, A., 2009. Wealth holdings and portfolio allocation of the elderly: the role of marital history. J. Fam. Econ. Iss. 30 (1), 90–108.
Ulph, D.T., 1988. A general non-cooperative Nash model of household consumption behaviour. University of Bristol Working Paper.
UN, 1995. Report of the Fourth World Conference on Women. United Nations, New York, NY.
U.N., 2005. Guide to Producing Statistics on Time-Use. Measuring Paid and Unpaid Work. UN Department of economic and social affairs, Statistics division, New York.
UN Statistical Commission, 1993. Report on the Twenty-seventh Session. Economic and Social Council Official Records, Supplement No. 6. United Nations, New York, NY.
Van Praag, B.M.S., Warnaar, M.F., 1997. The cost of children and the use of demographic variables in consumer demand. In: Rosenzweig, M.R., Stark, O. (Eds.), Handbook of Population and Family Economics. Elsevier, Amsterdam, pp. 241–274 (Chapter 6).
Van Rooij, M., Lusardi, A., Alessie, R., 2011. Financial literacy and stock market participation. J. Financ. Econ. 101 (2), 449–472.
Vella, F., 1998. Models with sample selection bias: a survey. J. Hum. Resour. 33 (1), 127–169.
Verheul, I., Thurik, R., Grilo, I., van der Zwan, P., 2011. Explaining preferences and actual involvement in self-employment: gender and the entrepreneurial personality. J. Econ. Psychol. 33 (2), 325–341.
Viitanen, T., 2014. The motherhood wage gap in the UK over the life cycle. Rev. Econ. Househ. 12 (2), 259–276.
Villeval, M.C., 2012. Steady, ready, compete. Science 335 (6068), 544–545.
Vogler, C., 1998. Money in the household: some underlying issues of power. Sociol. Rev. 46 (4), 687–713.
Vogler, C., Brockmann, M., Wiggins, R.D., 2006. Intimate relationships and changing patterns of money management at the beginning of the twenty-first century. Br. J. Sociol. 57 (3), 455–482.
Vogler, C., Lyonette, C., Wiggins, R.D., 2008. Money, power and spending decisions in intimate relationships. Sociol. Rev. 56 (1), 117–143.
Wagner, J., 2007. What a difference a Y makes—female and male nascent entrepreneurs in Germany. Small Bus. Econ. 28, 1–21. http://dx.doi.org/10.1007/s11187-005-0259-x.
Waldfogel, J., 1998a. Understanding the "family gap" in pay for women with children. J. Econ. Perspect. 12 (1), 137–156.
Waldfogel, Jane, 1998b. The family gap for young women in the United States and Britain: can maternity leave make a difference? J. Labor Econ. 16 (3), 505–545.
Ward-Batts, J., 2008. Out of the wallet and into the purse—using micro data to test income pooling. J. Hum. Resour. 43 (2), 323–351.
Warren, T., 2006. Moving beyond the gender wealth gap: on gender, class, ethnicity, and wealth inequalities in the United Kingdom. Fem. Econ. 12 (1–2), 195–219.
Weichselbaumer, D., 2004. Is it sex or personality? The impact of sex stereotypes on discrimination in applicant selection. East. Econ. J. 30 (2), 159–186.
Weichselbaumer, D., Winter-Ebmer, R., 2005. A meta-analysis of the international gender wage gap. J. Econ. Survey 19 (3), 479–511.
Weinberg, B., 2000. Computer use and the demand for female workers. Ind. Labor Relat. Rev. 53 (2), 2290–3008.
Weinberger, C.J., 1998. Race and gender wage gaps in the market for recent college graduates. Ind. Relat. 37 (1), 67–84.
Weinberger, C.J., 1999. Mathematical college majors and the gender gap in wages. Ind. Relat. 38 (3), 407–413.

West, C., Zimmerman, D.H., 1987. Doing gender. Gend. Soc. 1 (2), 125–151.
Wiepking, P., Maas, I., 2005. Gender differences in poverty: a cross-national study. Eur. Socio. Rev. 21 (3), 187–200.
Williams, D.R., 2012. Gender discrimination and self-employment dynamics in Europe. J. Socio-Econ. 41 (2), 153–158.
Wilmoth, J., Koso, G., 2002. Does marital history matter? Marital status and wealth outcomes among pre-retirement adults. J. Marriage Fam. 64 (1), 254–268.
Winter-Ebmer, R., Zweimüller, J., 1997. Unequal assignment and unequal promotion in job ladders. J. Labor Econ. 15 (1), 43–71.
Woolley, F., 1988. A non cooperative model of family decision making. LSE Working Paper tidi/125.
Woolley, F., 1996. Getting the better of Becker. Fem. Econ. 1, 114–120.
Woolley, F., 2003. Control over money in marriage. In: Grossbard-Shechtman, S.A. (Ed.), Marriage and the Economy. Cambridge University Press, Cambridge, pp. 105–128.
Woolley, F., Marshall, J., 1994. Measuring inequality within the household. Rev. Income Wealth 40 (4), 415–432.
Wozniak, D., Harbaugh, W.T., Mayr, U., 2010. Choices about competition: differences by gender and hormonal fluctuations, and the role of relative performance feedback. Working Paper, MPRA.
Wrohlich, K., 2004. Child care costs and mothers' labor supply: an empirical analysis for Germany. DIW Discussion Paper No. 412.
Wu, A.Y., Karamcheva, N.S., Munnell, A.H., Purcell, P.J., 2013. How do trends in women's labor force activity and marriage patterns affect social security replacement rates? Soc. Secur. Bull. 73 (4), 1–24.
Yamokoski, A., Keister, L.A., 2006. The wealth of single women: marital status and parenthood in the asset accumulation of young baby boomers in the United States. Fem. Econ. 12 (1–2), 167–194.
Yodanis, C., Lauer, S., 2007. Economic inequality in and outside of marriage: individual resources and institutional context. Eur. Sociol. Rev. 23 (5), 573–583.
Young, M., 1952. Distribution of income within the family. Br. J. Sociol. 3 (4), 305–321.
Young, M.C., Wallace, J.E., 2009. Family responsibilities, productivity, and earnings: a study of gender differences among Canadian Lawyers. J. Fam. Econ. Iss. 30, 305–319.
Zagorsky, J., 2003. Husbands' and wives' view of the family finance. J. Socio-Econ. 32(2).
Zelizer, V.A., 1989. The social meaning of money: special monies. Am. J. Sociol. 92 (2), 342–377.
Zelizer, V.A., 1994. The creation of domestic currencies. Am. Econ. Rev. 84 (2), 138–142.
Zick, C.D., Bryant, W.K., Srisukhumbowornchai, S., 2008. Does housework matter anymore? The shifting impact of housework on economic inequality. Rev. Econ. Household 6 (1), 1–28.
Zweimüller, M., Winter Ebmer, R., Weichselbaumer, D., 2008. Market orientation and gender wage gaps: an international study. Kyklos 61 (4), 615–635.

CHAPTER 13

Attitudes to Income Inequality: Experimental and Survey Evidence

Andrew E. Clark*, Conchita D'Ambrosio†
*Paris School of Economics—CNRS, Paris, France
†INSIDE, Université du Luxembourg, Walferdange, Luxembourg

Contents

Abstract

We review the survey and experimental findings in the literature on attitudes to income inequality. We interpret the latter as any disparity in incomes between individuals. We classify these findings into two broad types of individual attitudes toward the income distribution in a society: the normative and the comparative view. The first can be thought of as the individual's disinterested evaluation of income inequality; on the contrary, the second view reflects self-interest, as individuals' inequality attitudes depend not only on how much income they receive but also on how much they receive compared to others. We conclude with a number of extensions, outstanding issues, and suggestions for future research.

Handbook of Income Distribution, Volume 2A
ISSN 1574-0056, http://dx.doi.org/10.1016/B978-0-444-59428-0.00014-X

Keywords

Attitudes, Distribution, Experiments, Income inequality, Life satisfaction, Reference groups

JEL Classification Codes

C91, D31, D63, I31

13.1. INTRODUCTION

A number of areas of research in economics might sometimes be thought not to pass the "So what?" test: Do we really care about this issue? This would not seem to be the case for inequality, which looks like it passes the test with flying colors. Income inequality might be thought to occupy something like the same kind of place in the economic Pantheon as unemployment: It is almost taken as an axiom that it is a bad thing.

Given this sense of unanimity, it might seem to be churlish in the extreme to want to write a chapter about individuals' attitudes to inequality; surely they are negative aren't they? We believe that the situation is not quite as simple as might be imagined. First, we have to ask the rather fundamental question of what we mean when we talk about income inequality, and then why would we expect any measure of such inequality to be correlated with individual well-being. Following on from this setting-out of the scene, there are a number of open questions. Is inequality equally bad for everyone? And on an extremely practical level, how can we tell? Last, the term *inequality* is used perhaps rather loosely in the empirical literature. It is of interest to ask which measures of the distribution of income are the most important (to individuals) in this context: Is it (as is commonly assumed) the Gini coefficient, or rather something else? As we will discuss later, recent work using experimental and survey methods has allowed considerable progress to be made in answering some of these questions.

To set the stage, we first ask under which circumstances others' incomes should affect our own well-being.[1] We use the term *income inequality* to refer to any disparities in incomes between individuals (i.e., there is income inequality when some individuals have different incomes than others). As opposed to many of the other variables that have been related to individual well-being, the distribution of income does not exist at the individual level: income inequality is rather measured only at an aggregate, often societal, level. The key axiom in the measurement of inequality is the Pigou–Dalton principle of transfers, according to which inequality increases whenever a transfer of income from a poorer to a richer individual takes place.

[1] We limit ourselves here to discussion of individuals' evaluations of the inequality of income. Inequality in the distribution of other variables is of course of interest as well, including that of subjective well-being (as in Clark et al., 2014).

We believe that people do indeed have preferences over inequality. It is helpful to consider two broad types of individual attitudes to the distribution of income in a society. The first can be thought of as the individual's disinterested evaluation of income inequality: If I see two distributions of income in some society, which do I believe is better? We will call this the normative evaluation of inequality.

In addition to this disinterested reaction to income inequality, the individuals who we analyze when we carry out experimental or survey analysis do actually live in the society in question: Their own income then forms part of the income distribution in which we are interested. This second inequality effect is at the individual level: income inequality will directly impact both the absolute income that individuals receive, and how much richer and poorer they are compared to others. The attitude to inequality here is not disinterested but rather self-interested, with the additional assumption that individuals care not only about how much income they receive but also about how much they receive compared to others. We will call this the comparative evaluation of inequality.

The effect of the distribution of income on individual well-being will likely run through both of these channels. Even though income inequality as such in a society is not an individual-level concept, any distribution of income will have individual-level effects due to the way in which it changes the individual's own income and their standing with respect to those who are richer and poorer, as will be discussed later.

In the context of relative standing or comparisons, individual attitudes to inequality will depend critically on the *reference group* that the individual has in mind. This term was first used by Hyman (1942) in work on the evaluation of the rankings that individuals assign to themselves and refers to the group or individuals to which or with whom they compare themselves for the purpose of self-appraisal. The term has subsequently been refined and expanded in numerous contributions across the social sciences, with various definitions of the term now being proposed. Kelley (1965) distinguished between two roles that any such reference group can play and hence proposed separate definitions of the *comparative* and *normative* reference groups.

The first of these, the *comparative* reference group, is in the spirit of the original interpretation given by Hyman, whereby the reference group acts as the standard of comparison for self-appraisal. The *normative* reference group is the source of norms, attitudes, and values of the individuals concerned. Both groups can be further distinguished according to whether the individual in question *is* or *is not a member* of the reference group. Reinterpreting Shibutani's (1955) proposed conception of the terms, a comparative reference group is the point of comparison allowing the individual's own status to be calculated when the individual is part of the group (as in Hyman). However, the individual need not (yet) be part of the reference group. When the individual is not part of the group, but aspires to be, the reference group acts as a relative aspiration, that is, as the group of which the individual desires to be a member. A normative reference group is that whose perspectives constitute the frame of reference for the individual, and again a distinction between membership and nonmembership can be effected. In the latter case, individuals may adopt the behavior of the group as a result of anticipatory socialization (see Merton and Kitt, 1950).

Regarding the subject matter of this chapter, the reaction of an individual to income inequality will depend on both the role assumed by the reference group and membership status in the group. In a comparative reference group, of which the individual is a member, individual well-being is commonly assumed to be negatively affected by those who earn more than the individual, but positively affected by those who earn less. We say that the individual experiences *relative deprivation* from the income gaps with respect to those who are richer than she is in the reference group, but *relative satisfaction* from the income gaps with respect to those who are poorer. Both *relative deprivation* and *relative satisfaction* will very likely depend on the degree of income inequality within the reference group.

Comparative reference groups may also matter even if the individual is not currently a member of the group. If the individual aspires to be part of the group in question, then comparisons with respect to richer individuals in the group may give rise to positive feelings, as the individual anticipates being as rich as the group members once they join the group. This idea of a comparative reference group to which the individual aspires is akin to that of the tunnel effect in Hirschman (1973), which will be referred to in Section 13.2.1.

The rationale behind this comparative view of reference groups is that one's own position relative to others matters. We do not imagine that this is the only way in which others' outcomes may be viewed by the individual. It is very likely indeed that some groups will not be considered comparatively, but instead viewed with some kind of extended sympathy. The individuals to whom one compares and those for whom one feels sympathy are probably not going to be the same. As such, we may well see individuals whose position relative to their neighbors or work colleagues is paramount, but who at the same time vote for social programs for those in need or give money to international charities. Here individuals have a preference for making some others better off. We will explore this idea of empathy or altruism a little more in Section 13.4.4.

As opposed to the comparative view of reference groups, inequality in the normative view of reference groups is evaluated by the individual irrespective of where she appears in the distribution, or even irrespective of whether she appears in it at all. Concretely, a given distribution of income will be evaluated in the same way by an individual regardless of whether she is in the top or bottom quartile of the distribution, so that there is no role for comparisons to the richer and poorer in the normative reference group. Equally, we can all now have a normative opinion about the distribution of income in our own countries in the nineteenth century, even though we do not appear in that distribution ourselves. The normative evaluation of an income distribution can also be thought of as a mirror of preferences over inequality under the veil of ignorance (where the individual does not know where she will eventually be situated in the distribution).[2]

Both the normative and comparative views of income inequality will likely depend on how the distribution of income came about. We expect individuals to be more tolerant of the income gaps that result from effort than those that come about by luck. We will consider some of the work on the fairness of the income distribution further in Section 13.4.3.

The remainder of this chapter is organized as follows. Section 13.2 considers empirical evidence for an impact of income inequality in the context of a comparative reference group. We appeal to two different ways via which we can evaluate whether income inequality does indeed reduce the well-being the individuals who are exposed to it. The first approach relies on various measures of subjective well-being as proxies for individual utility: these are used to establish whether income disparities are indeed significantly associated with measures of individual well-being (such as happiness or life satisfaction).[3] The second is to see whether individuals behave as if they wish to avoid income inequality. This is tantamount to a revealed-preference argument. As it is anything but obvious to obtain clean measures of behavior and match these to income inequality in the field, we turn to experimental techniques in the laboratory to make progress here. Section 13.3 then follows the same structure, but this time with respect to the normative evaluation of income distributions. We propose a number of extensions, outstanding issues, and suggestions for future research in Section 13.4. Last, Section 13.5 concludes the discussion.

13.2. THE COMPARATIVE VIEW

When the reference group is viewed comparatively, individuals are not indifferent to others and compare to them to evaluate their own status in society.[4] If the individual is a member of this reference group, then higher incomes for others will reduce her well-being, whereas lower incomes have the opposite effect. Alternatively, if she is not in the reference group, but would like to be, then others' higher incomes will have a positive effect on her well-being. In both cases income disparities among others will be correlated with individual well-being.

[2] Considering only the level of one's own income, a richer-than-mean-income "impartial observer" will be more inequality averse when she is not involved in the distribution of income (this is the pure normative preference) as compared to the case in which she is present in it.

[3] This kind of subjective well-being literature has grown very quickly over the past couple of decades. As an example, three of the four top-cited articles published in the *Economic Journal* over the past 20 years have the word *happiness* in their title.

[4] Very generally, an individual's perception of inequality may depend on where she stands in the income distribution. An early contribution in this respect is Van Praag (1977).

We first consider evidence for the importance of such comparisons to others based on the measures of subjective well-being that are by now commonly available in many sources of survey data, before turning to the complementary work in experimental economics.

13.2.1 Subjective Well-Being and Others' Income

Arguably inspired by the salience of the Easterlin paradox[5] (Easterlin, 1974), and the increasing availability of information on various measures of subjective well-being in large-scale (including panel) data sets, there is by now quite a considerable stock of work on the relationship between income and well-being. One of the key questions in this literature has been "Does money buy happiness?" In standard economic theory, individual utility is not supposed to be affected by the behavior or income of others, unless these latter impose an externality on the individual.

In the context of the comparative reference group evoked earlier, however, the incomes of others in the reference group do indeed impose just such an externality. An increase in the income of others reduces the individual's well-being, through either greater relative deprivation or lower relative satisfaction (depending on whether the others whose income rises earn more or less than the individual in question), whereas analogously a reduction in others' income increases the individual's well-being.

There are any number of ways of attempting to show that individual well-being depends negatively on others' income. These were surveyed in Clark et al. (2008), and as such this chapter will only provide a shorter run-through of some of the relevant findings. Of course, comparisons need not be restricted to income and may well refer to comparisons of consumption, as initially suggested by Veblen (1949) and demonstrated empirically by, among others, Bloch et al. (2004), Brown et al. (2011), and Heffetz (2011). Comparisons could also cover leisure (Frijters and Leigh, 2008) or arguably almost any other observable economic attribute.

Some of the empirical work on the comparison of income has used a revealed-preference approach, in which observed measures of labor supply or consumption are argued to be more consistent with a relative utility function, in which either income or some consumption goods are compared to those of others in the reference group. A number of pieces of evidence along these lines can be found, for example, in Frank (1999), Layard (2005), and Schor (1992).

[5] This paradox is based on an opposition of the cross-section and time-series estimates of the relationship between subjective well-being and income. At any point in time, richer individuals are typically happier than poorer individuals. But as per capita GDP rises over time, Easterlin suggested that average subjective well-being remains flat in many countries. The extent to which subjective well-being is actually flat over time is the subject of quite heated debate (for example, Easterlin et al., 2010; Stevenson and Wolfers, 2008). The comparison of my income or consumption to that of others (or to myself in the past) has often been proposed as an explanation for this paradox.

It is always difficult to convince skeptics that any such correlations do indeed reflect spillover effects within the utility function, rather than learning, a hidden common factor within the reference group, or endogenous selection into the reference group. The tightest evidence in this respect may well come from natural experiments, in which either reference group income or consumption randomly changes. A small number of these experiments are described here.

Card et al. (2012) appeal not to expected outcomes but rather the revelation of information on others' earnings. The natural experiment here is a court decision that made the salary of any California state employee public knowledge. A local newspaper set up a website making it easy to find this information. Following this website launch, Card et al. informed a random subset of employees at three University of California campuses about the site. Some days later, all employees on the three campuses were surveyed. Comparing those in the treatment group (informed about the website) to others reveals the impact of information regarding others' salaries. The reference group in this work was defined as coworkers in the same occupational group (faculty vs. staff) and administrative unit in the university. Finding out about others' earnings should reduce the well-being of those who find themselves to be relatively less well-paid than others in the reference group and increase it if they find themselves to be better paid. The survey did indeed find lower job satisfaction for those with pay below the reference group median and a greater intention to look for a new job. The effect on both of these variables for those who were relatively well paid was insignificant. There is in addition some evidence of an actual quitting effect on those who were found to be in the bottom earnings quartile in the reference group.

Kuhn et al. (2011) consider observed large changes in close neighbors' incomes, which result from the design of the Dutch postcode lottery. Each week, this lottery randomly selects a postal code and allocates a prize of €12,500 per lottery ticket purchased within the postcode. In addition, one participating household in the winning postcode receives a new BMW. These postcodes are small, comprising on average about 20 households. Individuals who do not live in the winning postcode area, and those who do but did not buy a ticket, receive nothing. Households in winning postcodes were surveyed 6 months after the prize was won. One of the paper's key findings is that lottery nonparticipants in winning postcodes (who live next door to winners) are significantly more likely to have purchased a new car since the date of the lottery draw than are other nonparticipants, as if individuals do indeed compare their own car to that of their near neighbors.

A last example of a natural experiment is one in which comparisons to a reference position or an expectation (rather than comparison to other individuals) affect observable behavior (rather than subjective well-being). In New Jersey, police unions bargain over wages with their municipal employer, and in cases of dispute, an outside arbitrator has the final say. Mas (2006) found that the per capita number of crimes solved (cleared) is 12%

higher when the unions win their case compared to when they lose. He concluded that "the change in performance of police officers following an arbitration loss depends not only on the amount of the pay raise, but on the counteroffer that was demanded but never implemented as well" (p. 785).

Natural experiments of this kind are relatively rare. A great deal of work has instead appealed to survey data and modeled subjective well-being as a function of both the individual's own income and the income of a plausible reference group. This latter reference group is almost always imposed by the researcher as some measure of the income earned by those who are of the same age, sex, and education, for example, or who live in the same region, or (in the case of linked employer–employee data, as in Brown et al., 2008; Clark et al., 2009b) who work in the same firm. Direct information on who is in the individual's reference group in survey data is very rare (an exception is Clark and Senik, 2010).

Some of the by now large body of empirical literature is surveyed in Section 3.1 of Clark et al. (2008). For the income of "people like you," Clark and Oswald (1996) used the first wave of British Household Panel Study (BHPS) data to show that the estimated coefficients on income and others' income in a job-satisfaction equation are statistically equal and opposite, which is compatible with the Easterlin paradox. An early contribution by Cappelli and Sherer (1988) considered workers in the airline industry. The authors appealed to an occupational definition of others' earnings and showed that individual pay satisfaction is negatively correlated with an outside "market wage," which is average pay by occupation in other airlines. Ferrer-i-Carbonell (2005) related life satisfaction in the German Socio-Economic Panel (SOEP) to average income defined by sex, age, and education; Luttmer (2005) also considered life satisfaction, which is shown to be negatively correlated with average income by local area identified in a number of waves of the U.S. National Survey of Families and Households.

Instead of modeling reported subjective well-being as a function of own and others' income, an alternative is to ask how much income individuals need to attain a certain level of well-being. This is the method used in the Welfare Function of Income, associated with the Leyden school in the Netherlands. In this project, individuals are asked to assign income levels (per period) to a number of different verbal labels (such as "excellent," "good," "sufficient," and "bad"). It is then possible to estimate an individual lognormal Welfare Function of Income using the responses for each individual; this function shows how much income each individual needs to hit a certain level of well-being. The estimated means (μ_i) of these lognormal functions can then be used as the dependent variable in regressions seeking to explain which types of individuals require a higher level of income to be satisfied. The mean μ was found to be positively correlated with reference-group income (average income by age, education, and certain other individual or job characteristics); see Hagenaars (1986) and Van de Stadt et al. (1985). In other words, when the income of the reference group is higher, individuals need more money to attain a certain stated level of utility.

To date we have discussed empirical results that are consistent with a comparative reference group of which the individual is a member. The discussion in Section 13.1 revealed a possible counteracting effect when incomes rise in a comparative reference group to which the individual aspires, but of which she is not yet a member. Some work has indeed found that individual well-being is *positively* correlated with reference group income and has attempted to interpret this correlation in the light of aspirations and future outcomes. A positive correlation between my own well-being and others' income is consistent with Hirschman's tunnel effect, where others' earnings provide information about my own future prospects. In the terminology of Manski (2000), these are expectations interactions, where the individual updates their information set based on others' outcomes. The tunnel effect relates to the literature on the "prospect of upward mobility" (POUM), where both current and future income matter. This will be discussed further in Section 13.4.3.

Clark et al. (2009b) make the point that the estimated coefficient on others' earnings in a typical subjective well-being equation will likely mix together the comparison element (comprising relative deprivation and relative satisfaction, as discussed earlier) and the relative aspiration effect of the group to which the individual aspires. In the associated literature, this latter is often called an information or signal effect (whereas the former is called a jealousy or status effect). Positive subjective well-being effects from others' income are found, for example, in Senik (2004), Kingdon and Knight (2007), and Clark et al. (2009b). In each of these, the case can be made that the retained measure of others' income contains some element of my own likely future outcomes: An information or aspiration role for others' income is more likely the greater my probability of accession to the reference group in question. As will be discussed in Section 13.3.1, the inversion in the correlation between satisfaction and overall income inequality in Grosfeld and Senik (2010) in Poland can be interpreted in the light of such a tunnel effect. Individuals were initially happy with others' higher incomes (toward the top end of the income distribution), as this was thought to reflect their own future opportunities. Once it became clear that only relatively few people were actually going to be able to accede to these incomes, the correlation with satisfaction became more comparative, with a net negative effect in the later years of their sample.

Before describing the results of this literature any further, it is useful to set out the models of income comparisons formally. There is a set $N=\{1,\ldots,n\}$ of $n\geq 2$ individuals whose incomes are recorded in an income distribution $x=(x_1,\ldots,x_n)\in\mathbb{R}^n_+$, where $\mathbb{R}^n_+$ is the set of n-dimensional vectors with nonnegative components. The mean of x is $\lambda(x)$. For $x\in\mathbb{R}^n_+$, $B_i(x)=\{j\in N|x_j>x_i\}$ is the set of individuals with income greater than that of i, known as the better-off set; analogously, $W_i(x)=\{j\in N|x_j<x_i\}$ is the set of individuals who have an income that is lower than that of i, the worse-off set.

In the income-distribution literature, the most significant role of relative standing is in the determination of deprivation and satisfaction, which is related to inequality measurement as we will see later. As opposed to measures of income inequality, deprivation and satisfaction are defined at the individual level and aim to capture individuals' reactions when they compare their situation to that of others who have different levels of income (or of some other variable). Deprivation "involve(s) a comparison with the imagined situation of some other person or group. This other person or group is the 'reference group,' or more accurately the 'comparative reference group'" (Runciman, 1966, p. 11). In this literature, it is generally assumed that the reference group is the entire society.

The definition of relative deprivation adopted is the following: "We can roughly say that [a person] is relatively deprived of X when (i) he does not have X, (ii) he sees some other person or persons, which may include himself at some previous or expected time, as having X (whether or not this is or will be in fact the case), (iii) he wants X, and (iv) he sees it as feasible that he should have X" (Runciman, 1966, p. 10). When we consider income as the object of relative deprivation, which is the X in the preceding citation, then individual deprivation is simply the sum of the gaps between an individual's income and the incomes of all individuals richer than her.

Formally, Hey and Lambert (1980) specified the deprivation felt by someone with income x_i with respect to a person with income x_j as:

$$\begin{aligned} d_i(x) &= \left(x_j - x_i\right) \text{ if } x_i < x_j \\ &= 0 \text{ else} \end{aligned}.$$

In this case, as also suggested by Yitzhaki (1979), the deprivation function of an individual with income x_i is the sum of all the gaps to those in the better-off set divided by the number of individuals in the society:

$$D_i(x) = \sum_{j \in B_i(x)} \frac{x_j - x_i}{n}.$$

Aggregate deprivation, that is deprivation at a societal level, is then given by the average value of all of the individual deprivations. This aggregate deprivation turns out to the absolute Gini coefficient, which is given by the most popular index of income inequality (the Gini coefficient) multiplied by mean income.

Following on from these early contributions, Chakravarty (1997) proposed the inclusion of mean income in the measurement of individual deprivation. The latter now becomes the gap as a fraction of mean income, $d_i(x)/\lambda(x)$. This normalization is argued to be more appropriate for the comparison of the same society at different points in time, or different societies. When we use this formulation, aggregate deprivation is equal to the Gini coefficient, which is the absolute Gini index divided by mean income.

Analogously, income can be compared to those who are poorer than the individual in question (i.e., those who are in the worse-off set). This comparison yields the relative satisfaction function of an individual with income x_i, $S_i(x)$, given by:

$$S_i(x) = \sum_{j \in W_i(x)} \frac{x_i - x_j}{n}.$$

These measures of deprivation and satisfaction are called disadvantageous and advantageous inequality in Fehr and Schmidt's (1999) utility function. On this point Runciman (1966, p. 9) wrote: "If people have no reason to expect or hope for more than they can achieve, they will be less discontent with what they have, or even grateful simply to be able to hold on to it. But if, on the other hand, they have been led to see as a possible goal the relative prosperity of some more fortunate community with which they can directly compare themselves, then they will remain discontent with their lot until they have succeeded in catching up".

Although Fehr and Schmidt imagine that individuals are averse to both kinds of inequality, in the income-distribution literature it is most often implicitly assumed that individual well-being depends negatively on relative deprivation but positively on relative satisfaction. One of the main reasons for individuals not being inequality-averse, as will be set out in the following section, is that real income is not manna from heaven, and how that income comes about matters for individual attitudes.

This same concept of deprivation, which is at the core of the Gini coefficient, is also found in the literature of polarization (see Chapter 5). Deprivation is there called alienation. In general, alienation is assumed to be symmetric, whereas only the comparison to better-off individuals matters for deprivation. The interaction between alienation and identification is at the basis of the measure of polarization proposed by Esteban and Ray (1994). Bossert et al. (2007) reinterpret alienation and (the lack of) identification in terms of deprivation in a multivariate setting where functioning failures are analyzed. In this setting, individual deprivation is a multiple of the product of the share of agents with fewer functioning failures than the agent under consideration (the lack of identification) and the average of the functioning-failure differences between the individual and those who are better off (the alienation component).

The empirical subjective well-being literature described in this subsection has arguably made a key contribution in reminding social scientists (and maybe especially economists) that there are spillovers in individual income. The more you earn, the less happy I am, if you are in my reference group. Unless you are in a reference group to which I aspire, in which case my subjective well-being may well be higher (your position today provides me with an idea of what I can aspire to tomorrow).

The news is not only good, however. It can be argued that there are a number of drawbacks in this literature. In particular, the pertinent reference group is only a guess at who really matters in terms of the individual's own specific group that counts for income comparisons. In almost all cases, the best that we can do is use a series of likely reference groups and show that the effect of others' incomes seems to be consistent across them. An arguably useful piece of additional information comes from the identification of reference groups to which the individual aspires (for which there is an information or signal effect): We expect the correlation between individual subjective well-being and others' income in these groups to be less negative, or even positive. Even so, in both cases we can only guess at the correct reference group, with obvious implications for the accurate measurement of the relevant income gaps. As noted earlier in this subsection, we practically never ask individuals about their comparative reference group and have to our knowledge never asked about the reference group to which the individual aspires.

In the context of contributing to the analysis of relative deprivation and relative satisfaction described earlier, this literature has also not been an overwhelming success. Almost every paper here appeals to one single measure of the centrality of others' incomes, independent of whether the individual in question finds herself above or below that level. As such, there has been little attempt to distinguish relative deprivation from satisfaction.[6] Equally, knowing both my own income and the mean (or median) of my reference group income actually tells me fairly little about the gaps between me and others. Someone who has an income of 1000 euros above the mean or median reference-group income, say, can have widely varying values of relative deprivation and relative satisfaction.

The set of empirical subjective well-being work explicitly appealing to deprivation and satisfaction is not entirely empty. D'Ambrosio and Frick (2007) provided an empirical counterpart to the theoretical measures given earlier by exploring the relationship between self-reported income satisfaction and relative deprivation. Using panel data from the SOEP, they showed that subjective well-being depends more on a measure of relative deprivation than it does on absolute income because the correlation between income satisfaction and absolute income is 0.357, whereas that between satisfaction and relative deprivation is larger in absolute value at –0.439. As predicted by the income-distribution literature, the effect of relative deprivation on well-being is negative. This finding holds even after controlling for other influential determinants of well-being in a multivariate setting. Cojocaru (2014a) also estimated an individual well-being regression as a function of advantageous and disadvantageous inequality in the reference group, using 2006 data from the Life in Transition Survey (LiTS). Disadvantageous inequality is associated with lower life satisfaction, but advantageous inequality is not significantly so.

Bossert and D'Ambrosio (2007) introduced time as an additional dimension in the determination of the level of deprivation felt by an individual. They suggested that, as is usual, an individual's feeling of relative deprivation today depends on a comparison

[6] One exception, which arguably does fall into the group of survey work on satisfaction, is Loewenstein et al. (1989). Here individuals evaluate a series of hypothetical scenarios involving disputes between two people, where they are told to assume the role of one of the individuals and evaluate how satisfied they are with the final outcome in each situation. These satisfaction scores are shown to be related to both own and the other person's payoff. The correlation between satisfaction and advantageous inequality is much weaker than that with disadvantageous inequality.

with those who are better off today. They then proposed an additional consideration: The feeling of deprivation relative to someone who has a higher income today is more pronounced if this someone was *not* better off than the individual in question yesterday. In other words, relative deprivation is more keenly felt relative to those who, between yesterday and today, have passed the individual in question in the income distribution. Individual relative deprivation in this framework is then determined by the interaction of two components: the average gap between the individual's income and the incomes of all those who are richer than her (this is the traditional way of measuring deprivation), and a function of the number of people who were ranked below or equal in the previous period's distribution but who are now above the individual in question in the current distribution. A similar modification can be effected for the measurement of relative satisfaction, with the latter rising with the number of people that the individual has passed in the distribution between yesterday and today.

In a similar spirit to Bossert and D'Ambrosio (2007), D'Ambrosio and Frick (2012) proposed a utility function including dynamic-status considerations, which is tested on SOEP data. Individual well-being, measured in the SOEP by individual income or life satisfaction, depends at time t on four different elements: (1) the absolute component (i.e., the standard of living of the individual at time t); (2) the absolute dynamic component (i.e., how the individual's own income changed between $t-1$ and t); (3) the relative component, which is the individual's income at time t compared to others' incomes at time t; and (4) the relative dynamic component, which reveals how the result of the individual's income comparison in (3) changed between $t-1$ and t. This utility function is a generalization of that proposed by Fehr and Schmidt (1999), with the addition of individuals' income histories.[7]

This separation of income comparisons into those with respect to richer and poorer individuals, and explicitly distinguishing the others who have passed (or have been passed by) the individual in question, can be argued to shed some light on the debate regarding the potential status and signal effects of comparison income.

Individual well-being being negatively affected by comparisons to those who are permanently richer (and positively affected by comparisons to the permanently poorer) is completely in line with the standard empirical findings in the literature on relative income. At the same time, the presence of newly richer and poorer individuals can be argued to play the informational role described in Hirschman's (1973) tunnel effect. Someone who is today richer than me, but was yesterday poorer than me provides me with a positive signal about my own future prospects. And indeed in the empirical application, D'Ambrosio and Frick (2012) showed that individual satisfaction is positively

[7] Senik (2009) uses 2006 LiTS data, covering 28 post-transition countries (plus Turkey). She concluded that dynamic income comparisons (to oneself in the past) are more important than a number of other comparison benchmarks.

correlated with the income today of such people. Analogously, the income gap with respect to those who are now behind the individual but who were ahead of her reduces the individual's satisfaction, which is consistent with a negative signal that the individual could well be one of this group tomorrow. Finding such an effect in an advanced stable economy such as Germany is new and perhaps unexpected, in that previous work in the literature had rather underlined the relevance of the tunnel effect in societies that were either volatile or in earlier stages of economic development.

The broad conclusion from this work, which is by now far too voluminous to be listed in detail, is that others' incomes often do play a role in determining an individual's well-being. As the income of others to whom I compare rises, my well-being falls, but this status effect may be diminished or even entirely neutralized by a signal effect if what happens to others today informs me about what may happen to me in the future.

In general, however, the link between the formal models of income gaps (which are behind the measurement of inequality) and empirical work in the subjective well-being literature has been weak. The subjective well-being spillovers in society consist of a many-to-many mapping. As incomes in a society change, we need to know both who is affected by a movement in the income of individual *i* and who is in individual *i*'s reference group. We then have to identify the nature of the relationship between each pair: relative deprivation, relative satisfaction, or rather aspirations? Put in this light, it is obvious that we are asking a great deal of the information that is contained in standard surveys, all of which contain significant lacunae in this respect. To complement our understanding of how my well-being depends on my comparison to your income, we turn to experimental economics, where all the relevant parameters of the comparison process can arguably be controlled.

13.2.2 Experimental Economics

Experimentalists appeal to the notion of interdependence in preferences to explain the behavior of subjects who repeatedly violate game-theoretic predictions. Extensive surveys of work in this area can be found in Fehr and Schmidt (2003), Sobel (2005), and Camerer and Fehr (2006).

Interdependent preferences, that is, preferences that depend directly on the situation of others, were modeled formally for the first time in the theory of consumer demand. The phenomenon whereby individual utility functions depend on other people's income or consumption is known generically as the relative income hypothesis (Duesenberry, 1949). This can be further differentiated into "Keeping up with the Joneses," where the preference interaction with others depends on current consumption, and "Catching up with the Joneses," where it depends on lagged consumption. Leibenstein (1950) was the first to introduce demand functions that explicitly took into account the desire to be "in style," bandwagon and snob effects, and conspicuous consumption. Since then the literature has advanced to a considerable degree of sophistication, exploring the implications of such preferences on the theory of asset pricing (Abel, 1990; Campbell and Cochrane, 1999; Galí, 1994), Pareto optimality (Collard, 1975; Shall, 1972), the theory of optimal taxation (Abel, 2005; Aronsson and Johansson-Stenman, 2008; Boskin and Sheshinki, 1978; Dupor and Liu, 2003; Ljungqvist and Uhlig, 2000), the determination of work hours (Bell and Freeman, 2001; Bowles and Park, 2005), public spending (Ng, 1987), and the allocation of resources in general (Fershtman and Weiss, 1993), among others. A theory of social interactions has been proposed using varying formulations, where preferences are either defined over general consumption goods or an individual's identity. See Becker (1974) and Stigler and Becker (1974) for the first group and Akerlof and Kranton (2000) for the second. Sobel (2005) provides a thought-provoking discussion of the similarities and differences between these two strands of the literature.

Experimental work has made significant contributions to this area, in particular in considering the distribution of income across players, and distinguishing between doing better than others and doing worse than them.

13.2.2.1 Models of the Distribution of Income

The experimental economics literature fully incorporated distributional concerns into the utility function for the first time in Bolton (1991), with the modeling of *inequity* or *inequality aversion*. The two terms are very often used as synonyms in the literature to refer to the single phenomenon: that "people resist inequitable outcomes; i.e. the fact that they are willing to give up some material payoff to move in the direction of more equitable outcomes" as Fehr and Schmidt (1999, p. 819), to whom the definition of inequity aversion is due, put it.

The effect of inequality clearly results from some comparison being made to the reference group. On this point Fehr and Schmidt (1999, p. 819) continued by explaining that "Inequity aversion is self-centered if people do not care per se about inequality that exists among other people but are only interested in the fairness of their own material payoff relative to the payoff of others".

Fehr and Schmidt (1999) incorporated inequality into the individual utility function via the inclusion of all the pairs of the differences between the individual's own income and others' incomes. Bolton and Ockenfels (2000), who refined the earlier work of Bolton (1991), proposed an inequality-averse utility function that depends on the individual's own income and their share of the total income. The survey in Engelmann and Strobel (2007) compares these two approaches, together with that of Charness and Rabin (2002). Charness and Rabin's model is more related to social welfare than to inequality aversion and will not be analyzed in what follows: preferences in Charness and Rabin are a combination of the individual's own payoff and the payoff of the worst-off individual only.

Fehr and Schmidt (1999), who we henceforth call FS, proposed a utility function for individual i, $i=1,\ldots,n$, which depends on the individual's own outcome, and the gaps to those in the better-off set and the worse-off set, as defined in Section 13.2.1

$$U_i(x) = x_i + \alpha \sum_{j \in B_i(x)} \frac{x_j - x_i}{n} + \beta \sum_{j \in W_i(x)} \frac{x_i - x_j}{n} \tag{13.1}$$

where $\alpha \leq \beta \leq 0$ In this formulation, the utility of an individual depends positively on their own income, but negatively on both their levels of disadvantageous inequality (the gaps to those who earn more than them: the second term in Equation 13.1) and advantageous inequality (the gaps to those who earn less than them: the third term in Equation 13.1). According to Fehr and Schmidt, individuals dislike inequitable distributions. "They experience inequity if they are worse off in material terms than the other players in the experiment, and they also feel inequity if they are better off. (...) (H)owever, we assume that, in general, subjects suffer more from inequity that is to their material disadvantage than from inequity that is to their material advantage" (Fehr and Schmidt, 1999, p. 822). As such, α is larger in absolute terms than is β.

In the approach taken by Bolton and Ockenfels (2000), individuals are motivated by both their own pecuniary payoff and their relative payoff standing. They propose a theory of equity, reciprocity, and competition (ERC) in which the individual utility function is given by $U_i(x) = U_i\left(x_i, \frac{x_i}{\sum_{j=1}^{n} x_j}\right)$. The derivative of U_i with respect to the second argument is nonmonotonic, exhibiting a hump shape. This utility function satisfies a number of properties and, in a two-player game, with player i and j, one example of such an additively-separable utility function is

$$U_i(x) = a_i x_i + \frac{b_i}{2}\left(\frac{x_i}{x_i + x_j} - \frac{1}{2}\right)^2 \tag{13.2}$$

where $a_i \geq 0,\ \ b_i < 0$. In Equation (13.2), the utility of player i rises with her share of income when her share is under 50% and falls with her share when this share is over 50%.

In most experiments, these two models (FS and ERC) yield similar predictions. However, the predicted outcomes can differ for games where there are three or more players because ERC is not sensitive to all the inequalities in payoffs. In the ERC formulation, individuals want the average payoff of others to be as close as possible to their own but do not dislike the presence of richer and poorer individuals per se; in Fehr and Schmidt, individuals dislike inequality in all the outcomes. The experiment conducted in Engelmann and Strobel (2000) is designed to compare the performance of these two models: their results suggest that the formulation proposed by Fehr and Schmidt performs better than the ERC. A similar conclusion was reached by Dawes et al. (2007): Humans appear to be strongly motivated by egalitarian preferences.

The various contributions to the experimental literature measure inequality aversion via a number of alternative methods, which we will describe later. We believe that the appropriate term that should be used here is indeed inequality aversion, and not the original one proposed of inequity aversion. All the empirical contributions here are based on the assumption that the equality of payoffs is the fair, and hence equitable, outcome. But this need not necessarily be the case. If the distribution of income is not random, but depends (or is thought to depend) on individual effort or some other kind of meritworthy individual characteristic, the individual's view of what is equitable will depend on her own moral standards and the normative reference group. Opinions regarding what distribution of income is equitable will then very likely differ among subjects (see the discussion in Güth et al., 2009; Tyran and Sausgruber, 2006).

Experimental work has tested for the presence of inequality aversion and its consequences for economic outcomes in a number of different settings, such as ultimatum games, dictator games, dynamic bargaining games, public-good games with punishment, and redistribution games.[8]

13.2.2.2 Experimental Evidence from Ultimatum, Dictator and Dynamic-Bargaining Games

In the ultimatum game, some subjects, the proposers, are asked to suggest a division of a certain sum of money, say 100, between themselves and the other subjects, the responders. The proposer suggests a division, which the responder can either accept or reject. If the latter accepts the proposal, both the proposer and the receiver receive the money in accordance with the proposed division; if the responder refuses, neither player receives anything. Both the proposer and the respondent are fully aware of the rules of the game. The standard economic prediction based on subgame perfection is that the resulting outcomes will be very unequal: the proposer should make an offer of just over zero, and the responder should accept any positive offer that is made to them (as something is always better than nothing).

This prediction is not borne out by the behavior that is actually observed in the lab. The experimental results reveal a far more equal division of the pie, with responders frequently rejecting offers that are under 25% of the total sum (see Camerer, 2003; Levitt and List, 2007; see also Thaler, 1988, for a more comprehensive discussion of the general anomalies of these results). Bellemare et al. (2008) provide representative estimates of inequality aversion for the Dutch population. They found considerable differences between socioeconomic groups. Inequality aversion, in particular advantageous inequality, rises with age and falls with education level. Young and highly educated participants are one of the most selfish subgroups of the population under consideration. Fehr and

[8] More unconventional experiments have also been carried out showing preferences for fair redistributions (in experimental settings where effort can be controlled for) among Capuchin monkeys (Brosnan and de Waal, 2003) and 19-month-old infants (Sloane et al., 2012).

Schmidt (1999), in their survey of experimental results from the ultimatum game, noted that the vast majority of offers are consequently between 40% and 50% of the total sum, and no offers are below 20%. These results seem to hold regardless of the size of the sum that is to be divided, and in particular are also found in high-stakes games.

The second type of experiment used to reveal preferences over inequality is the dictator game. This is a simple variation of the ultimatum game, with the advantage of being nonstrategic. Here, as the name suggests, the proposer behaves like a dictator in proposing a split of the sum to be divided, with the responder having to accept the offer and thus having no decision to make. Experiments using the dictator game yield, as perhaps might be expected, distributions of income between the two players that are less egalitarian than those from the ultimatum game described earlier, with the proposer offering lower amounts. Even so, and despite the proposer running no risk of rejection, positive amounts of money are still offered. The survey of 616 such experiments in Engel (2011) concludes that dictators give on average 28.35% of the sum of money to be split to the responder, which is far from the self-interested economic prediction of no money being offered at all.

Abbink et al. (2009) also considered dictator games, but in the novel context of the destruction of others' income. This destruction is both negatively and positively framed. In the latter, individuals can decide to award their partner 50 points, and by doing so gain 10 points themselves. The decision not to make this award is analogous to the destruction of 50 of their partner's points at a cost of 10 points to themselves (and this is how the decision appears in the negative framing). Abbink et al. found destruction rates of about 25% with both framings. One surprising finding is that initially equal income distributions are actually more likely to be burnt, and the authors conclude as to the presence of a certain amount of equity aversion. One potential reading of this result is that, in their setup, the initially equal distribution is the only one from which the individual can gain rank by burning money (see their Table 1). We will return to the question of the rank comparisons of income in Section 13.4.2.

Last, in dynamic bargaining games, the evolution of bargaining proposals over time and the reasons that individuals provide for their behavior during the bargaining process can be examined jointly. In this framework, the experiments in Herreiner and Puppe (2010) show that Pareto-inferior solutions pertain due to the players' inequality aversion. For example, it is found that a majority (51%) of bargaining partners will reject the unequal payoff distribution of (46, 75) in favor of the Pareto-inferior equal split of (45, 45).

13.2.2.3 Public-Good Contributions and Punishment

In the public-good game, players are given an endowment and then secretly choose how much of this endowment they wish to put into the public pot (in order to finance the supposed public good, which will benefit everyone) and how much they would like to keep for themselves. Once the donation decisions have been taken by all players, the total sum of money in the public pot is multiplied by a factor of greater than one, and the resulting amount is evenly divided among all players. The Nash equilibrium in this game is for each player to contribute nothing to the public good. However, in experiments subjects are found to contribute an average of 40–60% of their endowment (Camerer and Fehr, 2004).

The public-good game can be refined by introducing a second stage in which information on others' contributions is provided, and players can punish each other. Introducing potential punishment in this second stage causes a sharp jump in cooperation in the first stage public-good game, as shown in Fehr and Gächter (2000). Masclet and Villeval (2008) assessed the role of inequality aversion in determining individuals' decisions to punish. They showed that individuals will punish others even when this punishment does not immediately affect the distribution of payoffs (in some situations the cost of one punishment point to the punisher is the same as the cost of this point to the target). Consistent with previous work, punishers are not primarily motivated by a desire to increase equality. Interindividual comparisons of outcomes do play a decisive role in the punishment decision in all treatments; the intensity of punishment is strongly correlated with the size of the difference in contributions and earnings between the punisher and the target. This result indicates that, irrespective of the willingness to directly reduce payoff differences, individuals may be willing to punish those whose decisions give rise to payoff differences, and that this inequality arouses emotions that trigger punishment. Punishment is shown to reduce inequality over time, as potential free-riders are incited to increase their contributions.

An open question in this literature is why individuals decide to spend their own resources to punish others. This decision could be self-centered, as today's punishment enhances my own future interests, or carried out altruistically in order to confer an advantage on my kin or group (see Van Veelen, 2012). Of course, any prosocial behavior can be self-interested if we include nonpecuniary moral preferences in the utility function (Levitt and List, 2007).

The sequential public-good game can be used to estimate separately the advantageous and disadvantageous inequality aversion suggested by Fehr and Schmidt (1999). In this game with two players, the first mover chooses his contribution to the public good under strategic uncertainty, as he does not know what the second mover will decide. The second mover does know what the first mover has decided and can choose to contribute either the same amount as the first mover or zero. Teyssier (2012) confirmed the theoretical predictions: First movers with greater risk aversion or disadvantageous inequality aversion contribute less to the public good than do others, and second movers with a sufficiently high degree of advantageous inequality aversion contribute more than do others. (For an analysis of risk aversion in the experimental literature see Section 13.3.2.).

Inequality aversion as in Fehr and Schmidt has been also applied to the analysis of the results of voting over redistribution. Although traditional economic models predict no redistribution, Tyran and Sausgruber (2006) showed that inequality aversion can predict the opposite result in their experiments, in which subjects have different endowments and decide how to redistribute from the rich to the poor by majority vote. On this point see also Farina and Grimalda (2011). In taxation games, Bolton and Ockenfels's ERC can predict the opposite allocations to those in Fehr and Schmidt, as shown by Engelmann and Strobel (2004), because the middle class would no longer be in favor of redistribution.

13.2.2.4 Deservingness: The Source of Income

One of the critiques of inequality aversion models and the experiments used to test them is that they often neglect the procedure that is behind the money to be allocated. Money appears here out of nowhere as "manna from heaven"; see, on this point, Bergh (2008) and Güth et al. (2009), among many others. In the majority of experiments, income is an allocation, so that having more than others is not seen as being deserved. However, in many real-world applications individuals likely believe that they earn more than others because they deserve to do so. As might be imagined, when income is considered to reflect effort rather than luck, the results do change. For example, Hoffman et al. (1994) reported that when the role of proposer in the ultimatum game is earned, rather than being randomly assigned, proposers offer less and respondents are more likely to accept unequal offers. Similar results are found in Cherry et al. (2002) when the asset of the dictators in the bargaining game is legitimate. We will return to this point in Section 13.4.3 when describing some evidence from the income-distribution literature on the fairness of outcomes. Another critique refers to the size of the stakes, with the suggestion that inequality aversion may be lower when the stakes are high. See on this point the discussion in Eckel and Gintis (2010), who concluded that this fact does not refute the theory but is rather a proof of the rationality of subjects who take the costs of their behavior into account.

A more general criticism of FS, which calls the scientific basis of their method into question, is contained in the various contributions of Shaked, and Binmore and Shaked. The details can be found in the January 2010 special issue "On the Methodology of Experimental Economics" of the *Journal of Economic Behavior & Organization*. This special issue includes the critique by Binmore and Shaked (2010a), the replies by Fehr and Schmidt (2010) and Eckel and Gintis (2010), and the rejoinder by Binmore and Shaked (2010b).

A novel test of the desire to change the income distribution and the provenance of the income in question appears in Zizzo and Oswald (2001). Rather than taking money from one person and giving it to another, participants in this experiment are allowed (at a cost to themselves) to destroy each other's earnings. This is the "negative framing" of the destruction described in Abbink et al. (2009) above. Participants played in groups of four. Each participant has the same amount of money to start with and can attempt to increase it by 10 rounds of betting on a number (1, 2, or 3) that is randomly chosen by a computer. A maximum amount per round can be wagered. This betting stage creates an unequal distribution of income. In the second stage, players can pay to burn each other's earnings, at a price to themselves of 0.01, 0.02, 0.05, and 0.25 of a money unit per money unit burnt.

Although the initial distribution of income is equal, two of the four players in each group are favored. These players can bet more than the others in each round of the betting stage, and they in addition receive a cash bonus between the betting and burning stages. This is public knowledge.

The results in Zizzo and Oswald show a remarkable amount of destruction. Just under two-thirds of players burned some money, and the average player had just shy of half of their earnings burned. The destruction rates here are higher than those in Abbink et al. (2009), which may well reflect that the average burning price here is lower. There is little evidence of a price elasticity of burning, except at the top burning-cost rate of 0.25. In the context of the current paper, richer players were burned more, but especially the two players who had received an unfair advantage were burned more.

13.2.2.5 Hypothetical Preferences and Neuro Evidence

Inequality aversion runs counter to the hypothesis that individuals are status seeking, as noted by Bolton and Ockenfels (2000, p. 172). The concern for relative standing is the focus of another set of contributions in experimental economics (see Alpizar et al., 2005; Johansson-Stenman et al., 2002; Solnick and Hemenway, 1998; Yamada and Sato, 2013). The approach here is to allow individuals to make choices over hypothetical states of the world to understand how important absolute and relative outcomes are to them. In income terms, these are couched in terms of own income and average societal income. The greater the importance of relative income, the more the individual will be willing to give up own income to achieve a better relative standing.

For example, in Solnick and Hemenway (1998), individuals are asked to choose between states *A* and *B*, as follows:

A. Your current yearly income is \$50,000; others earn \$25,000.

B. Your current yearly income is \$100,000; others earn \$200,000.

It is specified that "others" refers to the average of other people in the society and emphasized that "prices are what they are currently and prices (the purchasing power of money) are the same in States *A* and *B*."

The key in this hypothetical-choice literature is that respondents choose between one state in which they are better off in absolute terms and another in which they are better off compared with others. All of the cited papers find evidence of strong positional concerns over income, in that individuals report that they are willing to give up absolute income to gain status (choosing *A* over *B*). The percentage who exhibit "relative" preferences can be large: Half of the respondents said that they preferred to have 50% less real income but higher relative income (i.e., they preferred *A* to *B*; see Solnick and Hemenway, 1998, 2005).

Such choice experiments are easy to couch in terms of consumption or other life domains, rather than income, as well. The taste for relative standing in Solnick and Hemenway (1998) is found to be strongest for attractiveness and supervisor's praise and weakest for vacation time; in Alpizar et al. (2005) it is stronger for cars and housing and weaker for vacations and insurance. A useful extension in Corazzini et al. (2012) is to take the approach outside of only rich countries; in their work, respondents in high-income countries are more concerned by relative standing than are those in lower-income countries.

Most of these experiments have been conducted with students, which is the standard practice in experimental economics. Carlsson et al. (2007) is the first study that is based on a random sample of the population as a whole. Their results are comparable to those in Alpizar et al. (2005), who found that on average about half of the utility obtained from an additional dollar comes from relative concerns. Carlsson et al. (2007) reported that, on average, 45% of the utility increase from a small income increase arises from enjoying a higher relative income, a result that is halfway between 100% (corresponding to the hypothesis that only relative income matters) and 0% (where only absolute income matters).

A final set of experimental results comes from the recent NeuroEconomics literature. Fließbach et al. (2007) appealed to MRI techniques to measure the brain activity of pairs of individuals who carry out identical evaluation tasks in different scanners. If the individual succeeds in the task (remembering the number of blue dots on a previous screen, which they see for one and a half seconds), they obtain a monetary reward of a certain size, as indicated on their computer screen. The outcome of the other player (their success, and the amount won if the answer was correct) is shown at the same time. Fließbach and colleagues manipulated both the amount the individual won if correct and the amount the other player won to create a number of contrasting conditions. For example, in their conditions C6, C8, and C11, the individual always won 60 euros if his answer was correct (all participants were men), but the other player won, if correct, 120, 60, and 30 euros, respectively. One of each individual subject's many trials was randomly picked for payment after the end of the experiment.

The results show that relative incomes matter. Holding the subject's own earnings constant, the amount earned by the other player is significantly correlated with blood oxygenation level-dependent (BOLD) responses in the ventral striatum, one of the regions of the brain known to be involved in the processing of rewards. Wu et al. (2012) also found evidence of social comparisons in brain activity and suggested that it mostly appears in later cognitive appraisals and reappraisals, rather than in the initial evaluation stage. Recent follow-up work by Fließbach et al. (2012) repeated their 2007 experiment, but this time with both men and women, and distinguished between advantageous and disadvantageous inequality. Disadvantageous inequality is shown to have a much larger impact on brain activity in the ventral striatum than does advantageous inequality.[9] Dohmen et al. (2011) also used the same experiment and showed in a regression analysis that the effects of own and others' income on activation in the ventral striatum are equal and opposite (which was also true in the 2007 experiment). This holds for both men and women, although the estimated effect of both income variables is larger in size for men.

Somewhat similar in intent, although the experiment here consisted of individuals reading written reports on (fictitious) others who were superior or inferior to the respondent and the good or bad events that happened to them, is Takahashi et al. (2009).

Dawes et al. (2012) explicitly considered redistribution and brain activity. They considered individual decisions to pay a cost to change the distribution of income within a group, where this latter distribution was determined randomly. Redistribution was correlated with brain activation in an area known to reflect social preferences. In addition, this brain activation was shown to be correlated with survey measures of egalitarian preferences that were elicited outside of the scanner. Zaki and Mitchell (2011) showed that inequitable decision making (choosing to favor a smaller reward for oneself rather than a larger reward for the other player in a modified dictator game) is associated with brain activity in a region associated with subjective disutility. Last, Tricomi et al. (2010) explicitly addressed advantageous and disadvantageous inequality by randomly assigning individuals in pairs to be rich (with \$50) or poor (no dollars) after both received an initial allocation of \$30. Brain activity in areas known to be related to the valuation of stimuli was then measured via MRI as further transfers to both pairs were carried out. The results showed that the "poor" responded more strongly to transfers to themselves than to the other person, whereas the "rich" evaluated transfers to others more strongly than transfers to self. This is argued to show that individuals have social preferences over both advantageous and disadvantageous inequality.[10]

The discussion in the current section has shown that there is by now a considerable body of evidence consistent with individuals comparing their incomes with each other. Income is, in this sense, a social good. A certain amount of work has suggested something

[9] In a completely different setting, Cohn et al. (2014) also concluded that disadvantageous inequality matters more for effort decisions in a laboratory experiment than does advantageous inequality. Specifically, in a field experiment, individuals who reported that they were underpaid at an initial base wage increased their performance as the hourly wage rises; there was no such effect for those who reported being adequately paid or overpaid. Cohn et al. further showed that this distinction in the effort response to wages is only found for subjects who display positive reciprocity in a laboratory experiment.

[10] A novel contribution in the broad area of physiological reactions to income distribution is Falk et al. (2013). This paper first shows in an experimental setting that perceived wage unfairness (as in unmet expectations about the share of a reward to be received) is associated with measured individual heart-rate variability. It is also shows that the answer to a question on unfair pay in the 2009 wave of the SOEP is correlated with self-reported health outcomes and in particular with cardiovascular health.

of a loss aversion with respect to these comparisons, in that doing worse than others is more important in a well-being sense than doing better than others.

Any movement in the distribution of income will therefore affect societal well-being both directly, via changes in individuals' own incomes, and in a comparative manner, via the various gaps between individual incomes. Imagine a rise in inequality caused by an increase in some top incomes. Those who benefit from higher incomes will have higher well-being, both because they are richer and because their gaps to others have risen (although this effect may only be secondary). On the contrary, those whose incomes have not risen and who compare to the fortunate few who are richer are now relatively worse off, which reduces their well-being. The overall effect is a priori ambiguous.

Alternatively, inequality may fall due to a rise in the incomes of those at the bottom of the distribution (via an uptick in the minimum wage, say). Again, the well-being of those who benefit rises, both via greater own income and smaller gaps to the richer others. But the well-being of those who do not benefit falls as their advantageous gaps to the poorer are now smaller in size. If we continue to believe that this latter effect is of second order, then we may expect societal well-being to improve here.

Unfortunately, most of the changes in the distribution of income that we see are not this stylized. To make any kind of welfare statement, we need to know who compares to whom, how much the different kinds of income gaps matter, and how much relative income matters compared to absolute income. We have little reasonable hope of measuring these magnitudes with any degree of accuracy in existing data.

Even so, we do believe that the comparative reference group exists and represents one central constituent of attitudes toward inequality in an economy. The other main part of such attitudes comes from the normative view of inequality in the income distribution (as defined in the Introduction). Although there is a substantial amount of work devoted to the comparative reference group, it arguably turns out to be rather more difficult to evaluate normative attitudes toward inequality. It is to this question that we turn in Section 13.3. In this section we will also review some of the work that has tried to disentangle the various motivations behind individuals' actions.

13.3. THE NORMATIVE VIEW

In the normative view of the reference group, an individual evaluates the overall degree of income inequality in the reference group, but without making any comparisons to individuals who are richer or poorer than she is. Depending on the attitudes and social norms prevailing within a group, the individual can evaluate these income disparities as fair or unfair.

As in Section 13.2, regarding the comparative view of the reference group, there is evidence on the normative view of the reference group from both subjective well-being research and experimental analysis.

13.3.1 Inequality and Well-Being: What Do People Say?

We are interested in this chapter, as the title suggests, in individuals' attitudes toward or opinions about inequality. There are a number of ways in which these can be elicited, including direct questioning, experimental approaches, or inference from observed behaviors. In this subsection, we consider the contribution of "happiness economics," in which some measure of income inequality is related to the individual's self-reported well-being. In general, an equation similar to the following is estimated:

$$W_{ijt} = \alpha + \beta X_{it} + \gamma \text{Ineq}_{jt} + \varepsilon_{it}. \tag{13.3}$$

In this approach, we collect survey information on the subjective well-being of an individual i, living in some aggregate area j (where j is often, but not always, a country) at time t. This subjective well-being is related to a vector of standard demographic variables (age, sex, education, labor-force and marital statuses and almost always the individual's or the household's income) through the vector β. Of most interest to us here is the conditional correlation (i.e., controlling for all the variables in the vector X) between well-being and the aggregate measure of inequality in area j, Ineq_{jt}. The estimated value of the parameter γ shows us whether individuals, ceteris paribus, tick up or down their self-reported well-being scores in areas with higher or lower levels of income inequality.

The estimation of an equation like Equation (13.3) allows the "value of inequality," as it were, to be inferred from the empirical relationship between the observed inequality around the individual and their reported level of subjective well-being. This latter is most often measured by questions about the individual's happiness, life, and income satisfaction or some other measure of general psychological functioning. Multivariate regressions allow not only the sign of the conditional correlation between income inequality and subjective well-being to be established (γ shown earlier), but also the economic importance of any relationship that is identified (via the comparison of γ to some of the estimated β coefficients on other variables, such as income or unemployment).

This "happiness" approach to valuing public goods has now appeared a number of times in the subjective well-being literature. Some well-known pieces of work in this respect have considered inflation and unemployment (Di Tella et al., 2001), aircraft noise (Van Praag and Baarsma, 2005), and pollution (Luechinger, 2009), although there are by now many other applications.

Cross-section and panel data allow the happiness or satisfaction of tens or even hundreds of thousands of individuals to be measured. It is perhaps easy to get carried away by the sheer number of degrees of freedom here. Except that, as we suggest later, this is largely illusory: Although it is theoretically possible for each individual to be confronted with a different income distribution, the most common approach has been to take cross-country data, often repeated cross-section, and include the country-level Gini coefficient (or something else) on the right-hand side of a satisfaction regression. In this case, the effective number of degrees of freedom in the empirical estimation remains for the most part at the two-digit level.[11]

Although there are by now many thousands of empirical contributions across the social sciences that relate individual income to some measure of individual well-being, it remains true that only a small fraction of this existing work has considered any role for income inequality. Even so, it seems that the ease of access to large-scale data sets has led to relatively consistent growth of research in this area over time. A necessarily incomplete but hopefully somewhat-representative sample of some of the work that has been carried out in the area of income inequality and subjective well-being appears in Table 13.1. This table broadly reflects the growth in interest in the subject, but also considerable disparity in the estimated value of γ, as revealed by happiness data.

Perhaps the earliest contribution in economics is Morawetz et al. (1977), which contrasts two different Israeli communities and shows that the level of happiness is higher in the community with the more equal income distribution. Although interesting, the result essentially relies on two observations and does not control for all of the other factors that might differ between the two communities. A contribution that is more in the regression framework is an innovative article by Tomes (1986). This uses data (from the 1977 Quality of Life Survey) on individuals in approximately 200 Federal Electoral Districts in Canada. Matching in census data on income distribution, it is shown that the share of income received by the bottom 40% of the population is negatively correlated (at the 10% level) with both satisfaction and happiness for men. The same correlations are insignificant for women. Inequality is thus positively correlated with men's subjective well-being.

Hagerty (2000) is the first of a number of contributions to use U.S. General Social Survey (GSS) data. In his GSS sample from 1989 to 1996, maximum community income and the skew of community income are, respectively, negatively and positively correlated with happiness scores. Hagerty also used aggregate data from eight different countries to show that average happiness is lower in countries with wider income distributions. More recent work using the GSS has, however, come to a variety of results. Whereas Blanchflower and Oswald (2003) and Oishi et al. (2011) both concluded that there is a negative relationship between life satisfaction and income inequality, Alesina et al. (2004) and Di Tella and MacCulloch (2008) both found no significant relationships in GSS data. Alesina et al. (2004) is of interest here, as they explicitly compare long-run U.S. and European data, from the GSS (1972–1997) and Eurobarometer (1975–1992), respectively. Over the whole sample, inequality reduces reported subjective well-being among Europeans, but not Americans. The authors suggested greater (perceived) social mobility in the United States as one potential explanation of this difference.

[11] As one of the right-hand side variables in these kinds of regression is aggregated at a higher level than the dependent variable, the standard errors are underestimated and should be corrected as in Moulton (1990); it is not always clear that this correction is carried out in this literature.

Table 13.1 Income inequality and individual subjective well-being

Authors	Country	Data	Inequality measure	SWB measure	Inequality and SWB?
Morawetz et al. (1977)	Israel	Two different communities	Gini at community level	Happiness	Negative
Tomes (1986)	Canada	1977 Quality of Life Survey	Census data on share of income received by the bottom 40% in 200 Federal Electoral Districts	Satisfaction and happiness	Positive correlation for men
Hagerty (2000)	USA	GSS (1989–1996)	Maximum and skew of community income	Happiness	Negative
Hagerty (2000)	Cross-country	Eight countries	Gini	Happiness	Negative
Ball (2001)	Cross-country	1996 World Values Survey	Gini by country	Life satisfaction	Positive in raw data, positive and insignificant with controls
Blanchflower and Oswald (2003)	USA	20 Years of GSS	D5/D1 by state and year	Life satisfaction	Negative (but only significant for women, young, and the less educated)
Clark (2003)	UK	BHPS waves 1–11	Gini by region and year	Life satisfaction	Positive, especially for those who are more income mobile
Helliwell (2003)	Cross-country	WVS waves 1–3	Gini by country and year	Life satisfaction	No relation
Senik (2004)	Russia	5 Years of RLMS	Gini by region and year	Life satisfaction	No relation
Alesina et al. (2004)	USA	GSS (1972–1997)	Gini by year	Life satisfaction	No relation
Alesina et al. (2004)	Cross-country	Eurobarometer (1975–1992)	Gini by country and year	Life satisfaction	Negative
Graham and Felton (2006)	Cross-country	Latinobarómetro	Gini by country and year	Happiness	No relation
Schwarze and Härpfer (2007)	Germany	SOEP	Gini by region and year	Life satisfaction	Negative

Continued

Table 13.1 Income inequality and individual subjective well-being—cont'd

Authors	Country	Data	Inequality measure	SWB measure	Inequality and SWB?
Biancotti and D'Alessio (2008)	Cross-country	European Social Survey	Interquartile range by country	Happiness	Negative for those with more inclusive and moderate values
Bjørnskov et al. (2008)	Cross-country	WVS wave 3	Gini by country	Life satisfaction	No relation
Di Tella and MacCulloch (2008)	Cross-country	Eurobarometer and GSS (1975–1997)	Gini by country and year	Life satisfaction	No relation, but depends on other variables included in the regression
Ebert and Welsch (2009)	Cross-country	Eurobarometer (1978–1997)	Gini, Atkinson and hybrid measures by country and year	Life satisfaction	Negative
Knight et al. (2009)	China	2002 National household survey	Gini by county	Happiness	Positive
Berg and Veenhoven (2010)	Cross-country	World Database of Happiness (2000–2006)	Gini by country and year	Happiness	Slightly positive
Grosfeld and Senik (2010)	Poland	CBOS repeated cross-sections (1992–2005)	Gini by cross-section	Country satisfaction	Positive and then Negative as transition takes place (break in 1996). Positive effect only for right-wingers
Oshio and Kobayashi (2010)	Japan	Japanese General Social Survey (JGSS: 2000, 2003 and 2006)	Gini by prefecture and year	Happiness	Negative
Winkelmann and Winkelmann (2010)	Switzerland	Swiss Household Panel 2002	Gini by municipality/canton/region	Financial satisfaction	Negative

Oishi et al. (2011)	USA	GSS (1972–2008)	Gini by year	Happiness	Negative. Effect significant only for those in the bottom two quintiles of the income distribution. Effect is moderated by the perceived fairness of others, and whether the individual believes that others can be trusted
Verme (2011)	Cross-country	WVS waves 1–4	Gini by country and year	Life satisfaction	Negative
Van de Werfhorst and Salverda (2012)	Cross-country	ESS round 4	Gini by country	Happiness	Negative
Bjørnskov et al. (2013)	Cross-country	WVS waves 2–5	Gini by country and year	Life satisfaction	Effect more Positive the more the individual perceives society to be fair
Brodeur and Flèche (2013)	USA	BRFSS (2005–2010)	County-level percentage in poverty	Life satisfaction	Negative for all three of percentage of people of all ages in poverty, the percentage of related children age 5–17 in families in poverty and the percentage of people under age 18 in poverty in the county. Conditional on own income and neighborhood median income
Rozer and Kraaykamp (2013)	Cross-country	WVS waves 1–5	Gini by country and year	Average of Life satisfaction and happiness	Positive
Cojocaru (2014a)	Cross-country	LiTS wave 1	Gini by census enumeration area level	Life satisfaction	No relation

Schwarze and Härpfer (2007) calculated inequality in gross household income at the region and year level in 14 waves of German SOEP data. Life satisfaction is found to be negatively correlated with inequality (although a measure of income redistribution is not significant). Other work establishing a negative correlation between inequality and well-being includes Biancotti and D'Alessio (2008), Brodeur and Flèche (2013), Ebert and Welsch (2009), Oshio and Kobayashi (2010), Verme (2011), Van de Werfhorst and Salverda (2012), and Winkelmann and Winkelmann (2010), using data from a wide variety of different countries.

On the opposite side of the court, a number of contributions have instead concluded for a positive correlation. Along the same lines as the finding in Canadian data in Tomes (1986), Ball (2001) also found that happiness and inequality are positively correlated in raw data from the 1996 World Values Survey (WVS), although the introduction of a number control renders this positive correlation insignificant. The estimated value of γ in the first 11 waves of the British Household Panel Survey (BHPS) is positive (Clark, 2003), as is that in the first five waves of the WVS (Rozer and Kraaykamp, 2013). Last, in one of the relatively rare contributions entirely outside the OECD, Knight et al. (2009) found that county-level income inequality is positively correlated with happiness in the 2002 Chinese national household survey.

One recent intriguing contribution to this empirical debate comes from Grosfeld and Senik (2010). In contrast to a number of the contributions in Table 13.1, their identification is purely within and not between countries, as they consider data from Poland over its transition period. Using repeated CBOS cross-section data over the 1992–2005 period, they identified a turning point in the estimated relationship between inequality and subjective well-being. This correlation is positive and significant in the first years following transition, but then turns negative and significant. The break point that best fits this split in the data is 1996. The interpretation that the authors give is in terms of inequality first being regarded as providing opportunities for future higher incomes, which consequently turned into more negative comparative evaluations of disparities as it became clearer that not everyone would be able to benefit from any opportunities that this greater inequality promised.

As well as the sign and significance of the estimated effect, we are also interested in the size. Some of the work cited in Table 13.1 does contain explicit statements about marginal effects. For example, Tomes (1986) wrote that "an increase of 10% in the share of the poor reduces satisfaction by approximately 0.6 of a point. In order to maintain satisfaction unchanged, own income would have to be increased by \$4200 for every 1% increase in the share of the poor" (p. 435). This latter figure is larger than the annual income of 3860 Canadian dollars in his data set (although it should be noted that the confidence intervals around these estimates are quite large). Alesina et al. (2004) found that a one percentage-point rise in the Gini is compensated by a rise in annual income of 2950 dollars in the United States (8.7% of annual income) and 474 dollars in Europe (4.2% of annual income). The effect size in the SOEP in Schwarze and Härpfer (2007) seems more moderate: "If income inequality would be reduced by a half household income could be reduced by around 10% without changing life satisfaction" (p.244).

Although this kind of compensating differential is attractive in that it is easy to understand, it also obviously depends critically on the size of the estimated income coefficient in a subjective well-being equation. It is easy to believe that the coefficient on own income is actually an underestimate here, for standard endogeneity reasons, leading to trade-offs of income against inequality that are too high.

As an alternative, we consider the well-being effect of a one-point rise in the Gini coefficient, with the effect size being expressed as a percentage of the range of the subjective well-being measure. For example, the 0–10 life-satisfaction scale used in the SOEP has a range of 10; the corresponding 1–7 scale in the BHPS has a range of 6. It is not possible to calculate a standardized marginal effect using this metric across all of the work in Table 13.1. In the first instance, a number of the contributions here use ordered probit or ordered logit estimations, so that there are as many marginal effects as one minus the number of subjective well-being categories. Restricting ourselves to linear estimation techniques using the Gini, which yield significant estimates, cuts the sample down to five: Hagerty (2000), Schwarze and Härpfer (2007), Knight et al. (2009), Winkelmann and Winkelmann (2010), and Rozer and Kraaykamp (2013). These papers use five different data sets, with subjective well-being measured on a variety of scales.

Expressed as a percentage of the scale range, a 10% point change in the Gini coefficient mostly produces a movement in well-being of between 2% and 8% of the scale range (the exception being Schwarze and Härpfer, 2007, where the figure is smaller). In the SOEP, the standard deviation of life satisfaction is about 18% of scale range (1.79 for a scale of 0–10), with an analogous figure for the BHPS of 21% (1.29 for a 1–7 scale). A broad conclusion is that this very large movement in the Gini has an effect of between 0.1 and 0.4 of a standard deviation in life satisfaction. By way of comparison, the effect of unemployment on life satisfaction in the SOEP and the BHPS is somewhere around 6–10% of the scale range, or 0.3–0.5 of a life-satisfaction standard deviation.[12]

[12] It is arguably misleading to compare the size of the coefficient on inequality to that on individual unemployment. If half of the population are in the labor force, then a rise of 1% in the unemployment rate corresponds to one more person out of 200 in the population being unemployed rather than employed. Assuming that unemployment only affects the individuals who are unemployed (so that there are no spillovers) a one percentage point rise in the Gini index is roughly equal to a ten percentage point rise in the unemployment rate. For example, consider that subjective well-being is on a 1–10 scale, and the estimated coefficient on the Gini is –5: this ensures that a ten percentage point rise in the Gini will lead to a fall in predicted well-being of 0.5, which is 5% of the scale range (the midpoint of the figures mentioned in the text). If individual unemployment leads to an effect on individual well-being of 8% of the scale range (which is again the midpoint figure), then its estimated coefficient will be –0.8. A 1% rise in the Gini reduces well-being by 0.05 ($=0.01 \times 5$). A 1% rise in the unemployment rate will lead to a change in average well-being in the society by $-0.8/200=-0.004$. In this calculation, assuming no spillovers from the unemployed onto the nonunemployed, a rise in unemployment of over ten percentage points (12½ points, exactly) produces the same effect on societal well-being as a one percentage point rise in the Gini.

Some of the work on inequality and happiness here has explored the role of mediating variables or subgroup regressions to establish the subjective-groups for which the correlation with inequality is the largest to shed some light on the circumstances under which inequality affects subjective well-being. In the perhaps absence of a clear central tendency, it is arguably useful for policy purposes to know where and when inequality might be harmful in subjective well-being terms.

One of the best-known findings in this respect comes from Alesina et al. (2004): In Europe, inequality hurts the poor and left-wingers more (in the sense of having a greater negative effect on their well-being scores) than it does richer and right-wingers. This finding has recently been corroborated on more recent (2009–2010) Eurobarometer data by Vandendriessche (2012). Along the same lines, in Grosfeld and Senik (2010) the initial positive correlation between well-being and inequality was found only for right-wingers.

Other work has considered the mediating role of individual income. Oishi et al. (2011) found that the effect of inequality on happiness is negative and significant only for those in the bottom two quintiles of the income distribution. Schwarze and Härpfer (2007) found that only those in the first income tercile are negatively affected by post-government income inequality. In Clark (2003), the correlation between regional income inequality and individual well-being is more positive for individuals whose own income has been more mobile over time.

Oshio and Kobayashi (2010) carried out a number of tests of mediating variables and concluded that the correlation between happiness and inequality is more negative for women, the younger, those who have unstable positions on the labor market, and those who are politically in the center (rather than being progressive or conservative).

Some work has considered a mediating role for individual values, rather than observed demographic characteristics. In Biancotti and D'Alessio (2008), inequality has a more negative effect for individuals who report more inclusive and moderate values. Rozer and Kraaykamp (2013) found that the effect of Gini on well-being is more negative (actually less positive) for Europeans, those with more egalitarian norms (from a question on the relative preference for incomes being made more equal as opposed to needing larger income differences for incentive reasons), and those with greater levels of social and institutional trust. Last, as might be expected if the income distribution reveals information about the individual's own potential future position, in Ferrer-i-Carbonell and Ramos (2014) the effect of inequality is greater for those with higher (self-reported) measures of risk aversion in 1997–2007 SOEP data. The marginal effect of the Lander-Year Gini coefficient on life satisfaction is twice as negative for those with the highest risk-aversion score (on a 0–10 scale) as compared to the effect for those who report the modal score of 5.

One important individual value in the terms of this chapter, and one to which we shall return later, is the perceived fairness of the market system (i.e., the system that transforms individual inputs into individual outputs). In Oishi et al. (2011), the effect of inequality on happiness is moderated by the individual's perceived fairness of others and whether the individual believes that others can be trusted. Along the same lines, Bjørnskov et al. (2013) found that the perceived fairness of the income-generation process affects the association between income inequality and subjective well-being.

This burgeoning work on inequality and happiness has then revealed a number of intriguing findings. But perhaps one of the most striking aspects of Table 13.1 is the sheer variety of empirical correlations that have been uncovered. Is there any way of making sense of the variety of different estimated results here, or does sample variability rule the day (with as many positives as negatives as zeros)?

A first point, apparent from the fourth column of Table 13.1, is that there is no empirical agreement on the most appropriate measure of inequality. Although the majority of work refers to the Gini coefficient (a point to which we shall return in Section 13.4), it is also true that no consensus has been reached regarding the geographic level at which this coefficient should be evaluated.

Most of the empirical analysis has been carried out using data that contains only coarse-grained information on the distribution of income (i.e., at a very aggregated level, such as the country). Some work on British, Japanese, German, and Russian data has appealed to measures of inequality at the regional level (respectively: Clark, 2003; Oshio and Kobayashi, 2010; Schwarze and Härpfer, 2007; Senik, 2004). One of the few contributions to use large-scale data with more local-level inequality measures is Brodeur and Flèche (2013), who appeal to county-level information in the American BRFSS. Another is Winkelmann and Winkelmann (2010), who match in measures of inequality at all of the (in increasing order of size) municipality, region, and canton levels in the 2002 wave of the Swiss Household Panel. The research in Knight et al. (2009) combines more local-level measures of the distribution of income with data from a non-OECD country (China), finding a positive effect of the county-level Gini on respondents' happiness (see also Jiang et al., 2012).

One of the reasons why the degree of aggregation matters is that the Gini often moves only a little over time, a point made by Graham and Felton (2006), who noted that the Gini coefficient in Chile in the 2000s is not substantially different from that which pertained in the 1960s, despite the considerable social and economic changes that have taken place over the intervening period. Econometrically, it is then difficult to introduce both the Gini and country dummies into a regression, leading to the possibility that the Gini may be proxying for some other fixed country characteristic that is correlated with subjective well-being.

In general, this lack of variation in the measure of inequality does not help us to assuage the doubt that it is strongly correlated with some other variable that is important for happiness. For example, income inequality at the regional or country level could reflect industrial structure or the unemployment rate, both of which may well have independent effects on subjective well-being. Given a sufficient number of observations, it should be possible to tease out the independent contributions of inequality and other variables. But at the aggregate level it is anything but sure that sufficient observations are available. In general, the list of potentially important aggregate-level variables is often perilously close to the number of degrees of freedom in the analysis. In Di Tella and MacCulloch (2008), for example, income inequality attracts a negative but insignificant coefficient in their analysis of Eurobarometer and GSS data. They noted that this occurs "in part because there is some degree of co-linearity between the included variables. For example, if we do not include unemployment benefits, a variable that is highly correlated with inequality, we find that the coefficient on inequality becomes negative and significant" (p.36). Verme (2011) concurred that the lack of variability in survey measures of the Gini coefficient makes it particularly susceptible to multicollinearity with other aggregate-level variables (a problem he tackled via a number of robustness tests in which the other aggregate explanatory variables are dropped in turn).

An additional drawback to the empirical analysis of the relationship between individual well-being and aggregate income inequality is that it does not adequately distinguish between the comparative and normative aspects of the reference group. Even though some of the empirical analyses in Table 13.1 (although far from all) do introduce some measure of the mean of the income distribution into the analysis, they are unable almost by construction to calculate measures of relative deprivation and relative satisfaction from the survey data used. As such, any partial correlation between aggregate income inequality and individual subjective well-being very likely mixes together aspects of the comparative and normative reference groups, which perhaps explains the variety of estimated coefficients in Table 13.1.

Given the perhaps natural limits on the analysis of the relationship between aggregate inequality and individual subjective well-being, any evidence from this type of analysis will probably have to remain suggestive. This is arguably not the case for experimental work, where the reference group and the degree of inequality can be exactly manipulated, and it is to this that we now turn. Experimental work is of course not free of problems, in that what people say in a controlled setting may well differ from the way in which they would actually behave in reality, and their perceptions of inequality will likely be influenced by many factors. For a thorough discussion of these aspects and problems with experiments regarding social preferences, see Levitt and List (2007).

13.3.2 Experimental Economics

The experimental economics contributions to inequality aversion from the more aggregate perspective have appealed to two different approaches: (1) inequality and risk aversion with a parametric social welfare function; and (2) general social welfare functions. In the first of these, two types of experiments have been run. The first is similar to that adopted in the experiments on status or relative income discussed earlier in Section 13.2.2, that is the choice between alternative societies with different income distributions behind the veil of ignorance. The second type is based on the leaky-bucket experiment, which we introduce later.

Johansson-Stenman et al. (2002) carried out hypothetical-choice experiments. An individual's relative risk aversion is interpreted as the social inequality aversion from a utilitarian social welfare function's perspective. Inequality aversion is evaluated via individuals' choices between two types of society, from behind a veil of ignorance. Individuals are asked to choose the society that would be the best in terms of the well-being of their imaginary grandchild (in order for choices to be abstracted from the respondent's own circumstances and environment). The income distributions in the two societies, A and B, are uniform, and the respondent is told that their grandchild has an equal probability of receiving any income level within the range.

For example, Society A has a uniform income range of 10,000 to 50,000 Swedish kroner, whereas Society B has a uniform income range of 19,400 to 38,800 Swedish kroner. The student subjects in the experiment are told that prices are the same in the two societies, that there is no welfare state, and that there are no growth effects of the different income distributions.

An individual who is risk neutral will prefer Society A, in which expected income is higher. Someone who is indifferent between the two societies will have a relative risk-aversion parameter, η, that can be calculated by assuming a CRRA utility function[13] (see their Equation 5). In the example given earlier, indifference between societies A and B implies a value of η of 0.5; equally, an individual who prefers A (B) over B (A) will have a value of η of $<$ ($>$) 0.5. There are eight different conditions in their experiment. Society A always remains as described earlier, whereas there are eight society Bs, ordered such that indifference between A and B implies increasing risk aversion (see their Table 1). The higher is the value of η, the more income society is willing to give up to bring about a more egalitarian distribution of income, corresponding to a more concave social-welfare function.

The median value of inequality aversion in these experiments is in the interval between two and three. The respondents were fairly evenly distributed between the categories, with 43% of the respondents having inequality aversion of between one and five. Furthermore, a considerable number of respondents (17%) exhibited zero or negative inequality aversion. In addition, 19% of respondents exhibited extreme aversion compatible with the Rawlsian maxi-min strategy, which is the case of maximum aversion in the experiment. In a similar experimental setting, Carlsson et al. (2005) confirmed a median value of relative risk aversion of between two and three, and found a larger fraction of respondents (63%) with a value of relative risk aversion between one and five. In their experiment, 8% of respondents were found to be risk-lovers.[14]

[13] Such that $U = y^{1-\eta}/(1-\eta)$ if $\eta \neq 1$, and $U = \ln(y)$ if $\eta = 1$.

[14] It is notable that the values of the degree of inequality-aversion found in this experimental literature are far higher than those used in practice for the measurement of inequality: The U.S. Census Bureau uses a value of less than 1 (see http://www.census.gov/prod/2000pubs/p60-204.pdf), whereas the key inequality measures reported on the Luxemburg Income Study website as their "key figures" only use values of 0.5 and 1.

Some work in this area has tried to distinguish further between two types of inequality aversion: the first is the individual's level of risk aversion, as explained earlier, whereas the second is the individual willingness to pay to live in a more equal society. The estimation of individual inequality aversion only via risk aversion disregards any preferences that individuals may have regarding inequality per se.

To separate out these two attitudes, two types of experiments are carried out, one for each type of aversion. To this end, Carlsson et al. (2005) extended the analysis of Johansson-Stenman et al. (2002). The first experiment concerns the traditional imaginary grandchild, as described earlier, where the respondents do not know the position of their grandchildren, but only the income distribution and hence also the probability distribution in each society. In the second experiment, subjects choose between pairs of hypothetical societies with different income distributions, where the grandchild's income is known and is set equal to the mean income in the society. In other words, "In the first experiment individuals choose between hypothetical lotteries, where the outcomes determine their grandchildren's incomes in a given society. This experiment allows for the estimation of the individual's risk aversion in a setting where the level of social inequality is fixed. In the second experiment individuals choose between hypothetical societies with different income distributions, where the grandchildren's incomes are known and are always equal to the mean income in each society. This experiment enables us to estimate parameters of individual inequality aversion in a risk-free setting" (Carlsson et al., 2005, p.376).

In the second experiment, with a value of inequality aversion of zero, the individual is indifferent to income inequality; with a value of one, a 1% increase in own income yields as much utility as does a 1% fall in inequality. The median value of inequality aversion is found to be in the interval between 0.09 and 0.22, and most responses reflect positive inequality aversion. Only 7% of respondents appear to be inequality-lovers, in the sense that they are willing to sacrifice their own income to make society more unequal, whereas 6% are found to be extremely inequality-averse. Kroll and Davidovitz (2003) also found that subjects prefer more equal income distributions. However, when they had to give up part of their reward to shift to a more equal distribution, they chose not to do so.

Amiel et al. (1999) belongs to the second type of experiment in method (1), in which social inequality aversion is estimated via the *leaky-bucket* experiment. A sample of students were asked to indicate the amount of "lost money" that they were willing to accept for a transfer of money from a richer to a poorer individual, where this loss came about for example due to administrative costs. The median value of inequality aversion was estimated to be between 0.1 and 0.22, which is much lower than the existing estimates from the alternative approach, such as in Johansson-Stenman et al. (2002). However, the circumstances of the two experiments are very different, making a clear comparison of the results rather difficult.

That these large differences in the value of inequality aversion result from the different measurement techniques is confirmed by Pirttilä and Uusitalo (2010). The authors estimated inequality aversion using a questionnaire approach in a representative survey of Finns. The advantage of this questionnaire is that the same individual was asked questions based on two different measurement techniques: the leaky bucket and the preferred wage distribution under the veil of ignorance. The median value of the inequality aversion parameter from the leaky-bucket questions lay below 0.5. However, the results from the preferred distribution question gave a much higher value for inequality aversion, with the parameter being over 3. There are thus a considerable number of respondents who are willing to sacrifice the mean wage to bring about a more equal distribution of wages, but who at the same time are not willing to carry out costly transfers from richer to poorer.

Pirttilä and Uusitalo proposed a number of explanations for this rather radical difference in the results. One possibility is that people simply have different attitudes toward the implied efficiency–equity trade-off in different situations. The leaky-bucket question is specifically focused on redistribution, whereas the change in the wage distribution is a bargaining result. The two questions may also be measuring the same phenomenon but at a different scale. In addition, the leakage, that is the efficiency loss, is explicitly visible in the leaky-bucket question, whereas the respondent would have to calculate it in the wage-distribution question. Respondents may have had efficiency concerns in mind in the leaky-bucket question, and their preferences over efficiency could explain part of their unwillingness to support the transfer.

Pirttilä and Uusitalo also confirm the results in Beckman et al. (2004): the actual position of the respondent in the income distribution affects the answer given in the leaky-bucket experiment. As expected, support for this transfer is higher among the individuals who would benefit from it.

In the income-distribution literature the indices that are deemed appropriate to measure inequality are those that conform to the Lorenz dominance criterion. These indices fulfill four basic axioms: scale invariance, symmetry, the population principle, and the Pigou-Dalton transfer principle. For a recent survey of these properties and the dominance criteria see, among others, the excellent chapter in Chakravarty (2009). The first three properties are commonly assumed in the majority of indices of well-being; only the transfer principle, as we mentioned in the introduction, is at the heart of inequality measurement.

Attitudes toward inequality have been interpreted by some authors as being revealed by the reaction of (some relatively informed part of) the general public to these four basic properties. This is the contribution of the authors in group (2), where some general social welfare function is assumed but without any a priori functional form. The main question that is addressed in this part of the literature is what inequality seems to represent for the general public, and in particular whether these four basic axioms are reflected in individuals' views. The seminal book is this area is Amiel and Cowell (1999). Given that the defining concept for inequality measurement is the Pigou–Dalton transfer principle, we will discuss only those experimental results that cover this aspect of inequality.

In Amiel and Cowell (1992), the transfer principle is presented to respondents both as a numerical problem and verbally. In the former, they are asked to say which of two distributions of income are more unequal: $A=(1, 4, 7, 10, 13)$ versus $B=(1, 5, 6, 10, 13)$.

Verbally, they are asked to say what happens to inequality in the following scenario: "Suppose we transfer income from a person who has more income to a person who has less, without changing anyone else's income. After the transfer the person who formerly has more still has more."

Nearly two-thirds of the student sample in Amiel and Cowell (1992) did not think that inequality was lower in B than in A, whereas 40% did not agree that inequality would fall following the verbal scenario. The difference in these figures likely comes from individuals thinking of some kind of Robin Hood redistribution in the verbal case, whereas the actual numerical problem involves redistribution from the fairly poor to the even poorer. Amiel et al. (2012) examined many "flavors" or interpretations of the transfer problem. Only 21.6% of the sample are found to be in line with the researcher's standard view. A critique of the way in which some of these kinds of questions are asked is provided by Jancewicz (2012).[15]

Similar to Kroll and Davidovitz (2003) and Carlsson et al. (2005), Amiel and Cowell (2002), Gaertner and Namazie (2003), and Cowell and Cruces (2004), using method (2), examined the degree to which the principle of transfers is followed by people who evaluate inequality and risk. About 60% of respondents in the latter contribution viewed an equalizing transfer as inequality/risk reducing, and consistency in the risk version of the questionnaire was higher than consistency with the principle of transfers in the inequality version. This finding is confirmed by Gaertner and Namazie and Amiel and Cowell (2002), where the proportions of acceptance in the sample are 23% in the risk questionnaire and 17% for inequality.

Overall, individuals do have normative preferences over the distribution of income. It is, however, hard to argue that these are isolated in happiness regressions, as the latter are not able to separate out the comparative and normative components of attitudes to inequality. The experimental literature has been more successful in this respect, but even there the variety of different methods have produced quite a large range for the estimated value of inequality-aversion. Part of the problem here seems to be that the different methods make salient different preferences (such as risk aversion or preferences over efficiency). Another is that there are almost an infinite number of ways in which we can change the inequality of the income distribution, and preferences over taking money from the rich to give to the poor, and taking money from the middle or

[15] For example, the lack of a "Don't Know" response category, and there being no natural unit of account given for the figures in the numerical problem.

lower-middle class to give the poor may reasonably differ, even if the final impact on the Gini coefficient is the same.

13.4. OUTSTANDING ISSUES

This section discusses a number of issues that extend the existing literature on income gaps and income inequality described earlier.

13.4.1 Inequality and Other Outcome Variables

The discussion to date has considered individuals' relationship to others' incomes purely in the sense of "do they like it or not," whether that be revealed by survey information on subjective well-being or behavior in experiments. At the risk of opening a Pandora's box of other possible dependent variables, this is far from being the only outcome of interest. A number of other possible outcomes have been investigated across the social sciences. The following is a brief sample of some recent areas of research in this respect.

de Vries et al. (2011) tested the hypothesis that income inequality may produce individuals who are more competitive and less friendly toward others. These latter attitudes are captured by the Big Five personality factor of Agreeableness, which now appears in a number of surveys.[16] The regression analysis in de Vries et al. (2011) is based on almost 700,000 observations between 2001 and 2009 from an American web-based survey aimed at measuring personality. Agreeableness scores are significantly negatively correlated with state-level income inequality (as measured by the Gini coefficient on pretax household income matched in from the 2000 Census). This individual-level personality finding can be argued to be consistent with the considerable amount of existing evidence on aggregate inequality and measures of violent behavior. See Daly et al. (2001) for evidence on Canadian provincial-level murder rates and Macours (2011) for the role of district-level income inequality (over a period of income growth) in fueling civil conflict (as measured by mass abductions by Maoist rebels) in Nepal.

Loughnan et al. (2011) analyzed self-enhancement, which is the propensity to see yourself as being better than the average. They considered the relationship between self-enhancement and income inequality, with the argument that the gain from being better than others will be larger in more unequal societies. They administered a self-enhancement questionnaire to (mainly student) samples across 15 countries. In these questionnaires, respondents were asked about 20 different desirable characteristics. For each characteristic, they said whether they have more, the same, or less of it than the average student (or average person, in the nonstudent samples). They first showed that respondents on average think they have more of the characteristic than the average in

[16] As measured by the answers to questions on being interested in people, taking time out for others, and not being interested in other people's problems (this latter being reverse-coded).

14 out of 15 countries (the exception is Japan). They further demonstrated that self-enhancement is greater in countries with a higher Gini coefficient. This relationship is resistant to the introduction of a range of individual-level psychological variables.

In DeBruine et al. (2011), data from almost 5000 women aged 16–40 across 30 developed countries shows that women's preferences for facial masculinity are negatively correlated with a composite measure of country health: The value of masculinity as a proxy for developmental health is greater in countries where health is on average worse. Brooks et al. (2011) built on this work by noting that facial masculinity may also matter via the spread of the benefits that it confers. In the same way that a greater dollar return to higher rank in a golf tournament seems to lead to greater effort by players (Ehrenberg and Bognanno, 1990), any signal predicting competitive success is more valuable when rewards are more spread out. They hence match in data on the national Gini coefficient (from the United Nations Statistics Division) to DeBruine et al.'s original preference data. Their subsequent empirical analysis suggests that national income inequality is a better predictor of female preferences for facial masculinity than is national health.

Van de Werfhorst and Salverda (2012), in their introduction to a special issue of *Research in Social Stratification and Mobility*, suggested that income inequality at the national level is associated with a number of observable and attitudinal outcome variables. A number of the papers contained in this special issue go on to examine in detail the negative relationships between income inequality, on the one hand, and all of solidarity toward others, expressed support for democracy, and actual political participation.

Rothstein and Uslaner (2005) simultaneously estimated a measure of generalized trust and income inequality (the Gini coefficient). They concluded that inequality does indeed significantly diminish trust, whereas the estimated coefficient on trust in the income inequality equation is negative but insignificant. These findings have recently been critically reevaluated by Steijn and Lancee (2011), who specifically underlined the potential importance of non-Western countries with particularly high levels of income inequality and a confounding role of national wealth. Their regressions on Western country data (from the International Social Survey Programme, ISSP, and European Social Survey, ESS) show that the bivariate correlation between income inequality and trust is negative and significant, but that this relationship becomes insignificant in multivariate analysis once wealth is controlled for.

We are not necessarily arguing here that these additional potential attitudinal[17] correlates of income inequality are to be considered separately and in isolation. Rather, we think that they indeed represent some of the channels via which income inequality leads

[17] Moving beyond the individual level, we can also consider the attitudes expressed by other societal actors. Burgoon (2013) analyzes party position-taking in almost fifty years of annual data across 22 different countries. Net income inequality is positively and significantly associated with antiglobalization position taking.

through to overall well-being outcomes (and to those regarding individual health, on which there is a substantial literature that we have not covered here; see Chapter 17).

13.4.2 Other Measures of Different Aspects of the Distribution of Income

All our discussion of attitudes toward the distribution of income has been in terms of gaps to others in the reference group, in Section 13.2, and a normalized sum of all the gaps in society, as the Gini coefficient in Section 13.3. We have presumed that these are indeed the salient measures of others' income. But we do not know that for sure.[18]

Consider two log-normal distributions of income, where one is a horizontal displacement of the other, as in Figure 13.1. Which is the most unequal? If we are not in the income distribution then our (normative) evaluation of the dispersion in these two curves depends on which distribution measure we choose. Some measures of various aspects of the distribution of income are identical across the two; this is the case for the absolute Gini coefficient, the variance, the interquartile range, and the percentage of the population in relative poverty (as defined as income below 60% of the median, say). Other measures are

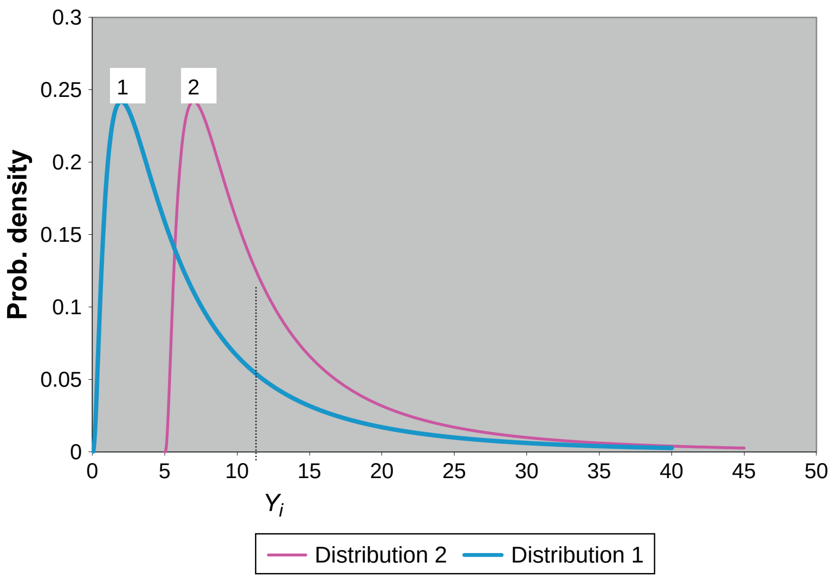


Figure 13.1 Two log-normal income distributions, with the same variance.

[18] A question that we do not address here is whether it is the pretax or post-tax income distribution that is correlated with subjective well-being. One reading is that it is the distribution of pretax income, which determines both well-being and preferences for redistribution, and this latter influences the actual tax system, which in turn determines the posttax distribution of income. At the practical level, not all empirical papers make clear whether their income measures are net or gross.

not the same in distributions 1 and 2: the percentage in absolute poverty, the relative Gini coefficient, and the D9/D1 or D5/D1 ratios.

If the individual making the evaluation is in the income distribution, then their evaluation will also depend on their own income position; this is the comparative evaluation. At an income of Y_i, an individual will feel more deprived in distribution 2 than in distribution 1; their relative deprivation will be higher (more people above them), and their satisfaction will be lower (fewer people beneath them).

We have considered the relationship between objective measures of inequality, such as the Gini, and subjective well-being. But do people actually know what the value of the regional or national Gini coefficient is? Individuals' perceptions of the degree of inequality around them may not be well reflected in the Gini coefficient, and equally they may believe the distribution of income to be different from what is actually measured in statistics.

Macunovich (2011) is an intriguing contribution using the fourth (2005) wave of the WVS. She analyzed not only the Gini coefficient, but also two measures of crowding at the bottom of the distribution: the ratio of the number of people who say that they are in the lowest income decile in the country to the number who say that they are in the highest, and the same ratio with respect to self-reported social class. Although the Gini coefficient continues to exhibit a positive correlation with both happiness and life satisfaction, the estimated coefficients on these two ratios are negative and often significant. This might be thought of as consistent with some of those at the bottom providing a negative signal (as in D'Ambrosio and Frick, 2012), or more generally with some negative externalities in society associated with poverty.

O'Connell (2004) used information on (the log of) the income received by the top quintile in a country to that received by the bottom quintile. This is shown to be negatively associated with life satisfaction in an aggregate-level analysis of 15 EU countries in Eurobarometer data over the 1995–1998 period.

Some of the work appearing in Table 13.1 uses a variety of measures of income distributions. Both Tomes (1986) and Brodeur and Flèche (2013) considered the bottom end of the distribution, with the former including the share of income earned by the bottom 40%, and the latter the county-level percentage in poverty according to three separate definitions.

In general, however, very few contributions here have tested different measures against each other in a beauty contest to see which one is the most salient correlate of subjective well-being. Ebert and Welsch (2009) is relatively unusual in this literature in that they did consider a wide class of inequality indices comprising the Atkinson and Gini family as special subclasses (see Ebert, 1988) and evaluated their effects on individual reported life satisfaction in 20 years of Eurobarometer data. As the self-reported income data in the Eurobarometer is not sufficiently good to allow detailed measures to be computed from within the data set, these latter are matched in from the

Luxembourg Income Study (LIS), which somewhat reduces the number of countries that can be used in the empirical analysis.[19]

Ebert and Welsch started by considering the Gini, Atkinson 0.5, and Atkinson 1.0 indices. All three of these are shown to be significantly correlated with life satisfaction in ordered probit regressions. They then considered generalizations of these indices. Their analysis of life satisfaction leads them to conclude that both rank and level inequality aversion matter, and that the overall degree of inequality aversion is larger than that implied by the standard measures applied in empirical analysis.[20]

A last point with respect to the question of "which measure of others' income" is that existing work has very much concentrated on cardinal measures of comparisons, as picked up by income gaps and Gini coefficients. Although there is likely some role for such comparisons, it also seems probable that individuals are rank-sensitive. Some previous work has considered the role of income rank in determining well-being. In Brown et al. (2008), income rank is shown to outperform average reference group income in three satisfaction equations (influence over the job, achievement, and supervisor's respect): see also Clark et al. (2009a) for economic satisfaction and Boyce et al. (2010) for life satisfaction. In the field experiment in Card et al. (2012), information on the individual's revealed rank in the income distribution was more important in determining their satisfaction than was the relative wage level. Clark et al. (2010) appeal to both survey and experimental evidence on the role of relative income in determining the level of effort that workers supply. In both types of data, the individual's rank in the income distribution is a more powerful determinant of their effort decision (as measured by the log-likelihood) than is the relation of the individual's own income to mean income in the reference group. Mujcic and Frijters (2013) came to the same conclusion in the analysis of hypothetical choice data from a sample of just over 1000 Australian students. Finally, Clark and Senik (2014) appealed to Chinese panel data from Guizhou province, in which all households in the village were interviewed. This complete data allows household rank in the village income distribution to be determined. Being at the top (top decile) or bottom (lowest 25%) of the income distribution seems to matter disproportionately for satisfaction with income.

It may also be the case that not all ranks are of equal importance, so that the correlation between income rank and subjective well-being is nonlinear. The experimental and survey results in Kuziemko et al. (2014) underline the importance of the aversion to being last in the distribution. Experimental subjects accept gambles, which may move them out

[19] Although not in the context of subjective well-being, Jancewicz (2014) provides an extremely interesting analysis of the criteria that individuals use to sort different income distributions into groups that have similar perceived levels of inequality.

[20] It would also be of great interest to evaluate the relationship between income polarization and individuals' reports of subjective well-being. We are not aware of any contributions in this respect.

of last place that they reject if anywhere else in the distribution. Equally, subjects randomly placed in second-to-last place in modified-dictator games are the most likely to give money to the person one rank above them instead of the person one rank below. One implication is that the relatively poor may oppose redistribution if it is especially targeted at those who are just beneath them in the income distribution. Survey data does indeed show that respondents who earn just above the minimum wage are those who are the most likely to oppose any rise in the minimum wage.

These kinds of rank comparisons are of great interest. They do imply a role for inequality in the determination of individual well-being in that, given own income, a mean-preserving spread of income in the reference group implies lower individual rank. However, at the societal level this will not matter. By construction, rank is zero-sum: My loss must be offset by others' gains. Unless we have heterogeneity in the taste for rank (as in Frank, 1985), the degree of inequality will here not affect the way in which income comparisons affect overall well-being.

13.4.3 Fairness and Preferences for Redistribution

The measures of income distribution used in the preceding literature have been objective: They measure what others in the society actually earn. This is of course not necessarily what individuals *believe* that others earn, and it may well be this latter, and its relation to what it is believed that others *should* earn, that is the most important for determining individual attitudes toward inequality.

Almost no one in our societies thinks that everyone should receive the same income. Incomes differ often for very good reasons, such as number of hours of work for example. In general, we can think of the causes of income distribution as being partitioned into factors for which the individual is responsible and those for which she is not (see Fleurbaey, 1995). These are respectively referred to as effort and circumstances in the literature on the equality of opportunity (see Chapter 4 of this volume for a survey in great detail).[21] Almås et al. (2011) proposed the measurement of a "responsibility-sensitive" fair income distribution. This is applied to 1986–2005 Norwegian data. They show that although the Gini index fell over this period, unfair income inequality actually rose. Further, the pretax unfair income Gini rose less than the posttax unfair income Gini, so that the tax system has become less profair.

An alternative approach to fairness, which does not require the explicit distinction of responsibility and nonresponsibility factors, is to explicitly ask individuals about how much they think others should earn. For example, the cross-country ISSP surveys have asked direct questions a number of times about perceived and fair distributions of

[21] We might expect inequality to be less acceptable when it occurs by chance, rather than from individual effort. In this context, it is noticeable that there is no particular push to redistribute from lottery winners. This may reflect that they on average already pay a tax by spending more on the lottery than they receive. In the UK National Lottery, for example, less than half of the money spent on tickets is won in prizes.

incomes. Each year the ISSP survey administers a number of core questions, as well as rotating modules on specific topics. These modules in 1987, 1992, 1999, and 2009 were on Social Inequality. Individuals were asked directly how much they thought that individuals in certain job types earned. For example, in the 1987 wave, variable v26 refers to the answer to the following question:

> *We would like to know what you think people in these jobs actually earn. Please write in how much you think they usually earn each year, before taxes. (Many people are not exactly sure about this, but your best guess will be close enough. This may be difficult, but it is important, so please try.). First, about how much do you think a bricklayer earns?*

Variable v27 refers to the answer to the same question, but now with respect to a doctor in general practice. The following nine questions then cover the income of a bank clerk, an owner of a small shop, the chairman of a large national company, a skilled worker in a factory, a farm worker, a secretary, a city bus driver, an unskilled factory worker, and a cabinet minister in a national government.

Last, individuals are asked a series of 11 questions covering the same occupations, but this time are asked to indicate what they think that these individuals *should* earn each year before taxes, regardless of what they do actually receive.

The same types of questions are repeated across the different Social Inequality modules, although by 2009 the questions only covered the five occupations of a doctor in general practice, the chairman of a large national company, a shop assistant, an unskilled factory worker, and a cabinet minister in a national government.

Similar kinds of questions have appeared in a number of other surveys, including the 2005 wave of the SOEP. It is also possible to ask these questions about actual and just rewards with respect to the individual herself, or regarding a hypothetical third person with a given set of demographic characteristics (see Jasso, 2007).

One application of the answers to these questions is to consider the responses that are given for occupations at the top and bottom end of the income distribution; for example, in the preceding ISSP questions, the incomes of the chairman of a large national company and an unskilled factory worker. The ratio of these two gives an indication of the income inequality that the respondent perceives. Along the same lines, a fairness index can be calculated as the ratio of the incomes that the individual believes that these two occupations should earn. The comparison of these two ratios gives an indication of how much of the gap in earnings that the individual perceives is considered to be fair.

An empirical application of this kind of approach can be found in Schneider (2012), who used German data from the 2006 wave of International Social Justice Project to consider the relationship between subjective well-being and income inequality. Instead of calculating a Gini coefficient from within the data set, or matching it in at some level from an external data source, she calculated a direct measure of the individual's perception of the fairness of the income distribution.

Using the responses to the questions about the perceived incomes (PI) and just incomes (JI) of a managing director (MD) and an unskilled worker, she calculated a measure of the overall legitimacy of income inequality as follows:

$$\text{legitimate inequality} = \ln\left[(PI_{MD}/PI_{unskilled})/(JI_{MD}/JI_{unskilled})\right].$$

Someone who believes that the current income distribution is just has a value of legitimate inequality of zero. Those who believe that the income gaps should be wider will have a negative value, whereas those who perceive some inequality of reward will have a positive value. This measure of legitimate inequality varies at the individual level, therefore providing far more potential explanatory power than the aggregate-level measures of income inequality that have been discussed so far.[22]

In her 2006 data, Schneider reported an average value of the first term in the square brackets, $PI_{MD}/PI_{unskilled}$ of around 644, with the average value of $JI_{MD}/JI_{unskilled}$ being slightly over 300. This yields a value of legitimate inequality of around 0.75. The individual level of the perception of inequality is shown to be negatively correlated with life satisfaction. This correlation is stronger for the higher than for the lower income groups.

Osberg and Smeeding (2006) appealed to these questions in the 1999 ISSP survey. However, instead of looking at the gaps with respect to the top and bottom occupations in the list, they considered the entire set of responses regarding perceived and just incomes. By assuming that there are equal numbers of individuals in each of the nine occupations, they could calculate Gini coefficients, both with respect to the income that the individual actually believes is earned and a "just" Gini coefficient for the income that she believed should pertain. They then calculated the ratio of these two Gini coefficients: a value of less than one implies that the individual believes that there should be less inequality than that which she believes exists.[23]

Most people are in favor of some leveling of incomes, whereas very few believe that all incomes should be the same. The average value of the ratio of the Gini coefficients is less than one in all of the 27 countries that appear in the 1999 ISSP. The average figure across all countries is 0.75. In some countries, such as the United States and Japan, this figure is around 0.8; in others such as Spain and Sweden, it is under 0.7. As well as cross-country differences in this measure of the desire to redistribute, there are systematic

[22] Legitimate inequality does differ across individuals, which is good. However, it is also potentially endogenous, with unhappy people thinking that people at the top of the income distribution earn more (or should earn less), for example.

[23] Blanchflower and Freeman (1997) are along the same lines. They used the 1987 and 1992 ISSP surveys and considered the standard deviation of the log response given across the different occupations. They did this both for the perceived and fair distributions (which they call the perceived and appropriate differentials). Their main result is that ex-Communist countries both perceive and consider appropriate tighter income distributions compared to Western countries, but that this gap fell sharply over the transition process (here between 1987 and 1992).

differences by individual characteristics. In particular, Osberg and Smeeding underlined the importance of age, education, and family income in this respect.

A more direct approach to perceived income inequality, and its relation to subjective well-being, was taken by Smyth and Qian (2008), who used Chinese data from a 2002 survey of 31 cities. In this data, individuals were asked directly about their perceptions of inequality of the income distribution, on a 1–5 ordered scale. These perceptions are shown to be correlated with individual happiness scores.[24] The sign of this correlation depends on the individual's own position in the income distribution. In particular, perceived inequality is negatively correlated with happiness for individuals who are in the bottom quintile of the income distribution, whereas this correlation is positive for those who are in the top quintile.

The importance of inputs and income inequality has also appeared in the experimental literature: It matters where income comes from. Abeler et al. (2010) found, in a gift-exchange game, that equal wages lead to systematically lower levels of effort being furnished by workers than when the firm can decide to pay workers differently. Their explanation is that workers do not want their wage–effort ratio to be lower than that of their coworkers, and they consequently work less hard. Clark et al. (2010) also found that the wages offered to other experimental participants in a gift-exchange game are negatively correlated with the effort furnished by the worker. Krawczyk (2010) found in experimental work that the equality of opportunity moderates the desire to redistribute.

Perhaps the most direct evidence of attitudes toward inequality comes from asking individuals whether they want to see less of it, by redistributing from the richer to the poorer. There is a considerable literature on the desire to redistribute (see Förster and Tóth, 2014). One of the first contributions is Persson and Tabellini (1994), who both proposed a theoretical model and presented some empirical results with respect to the median-voter theorem. The individuals here are purely concerned with their own self-interest and have no social preferences as such. The median here refers to the distribution of some variable, for example, income or skills (as measured by education). The individual's voting preferences will then depend on their own position in that distribution.

A second well-known contribution is Piketty (1995), who developed a theoretical model to explain why, in the long run, left-wing dynasties in the lower class are more supportive of redistributive policies, whereas right-wing dynasties in the upper-middle classes are less so. As in Persson and Tabellini (1994), individual income is here related to

[24] This correlation is arguably large in size. Happiness in their survey is on a 1–5 scale, as is the individual's perception of inequality. The overall partial correlation between happiness and fairness in their Table 2 is –0.09. As such, the effect of moving from the bottom to the top of the perceived income inequality scale has an effect of 0.36, which is 9% of the scale range. We cannot directly compare this figure to the correlation between subjective well-being and the Gini coefficient in Section 13.3.1, as we cannot map the seriousness of inequality to a particular Gini figure.

political opinion: those with higher incomes are more right-wing and less favorable to redistributive policies, whereas those with lower incomes are more likely to vote for left-wing parties and to be in favor of redistribution.[25]

It is not only the individual's situation today that counts, but also where she thinks she might end up tomorrow. The "prospect of upward mobility" (POUM) literature explicitly appeals to individuals' future prospects of social mobility. As such, own current income is not a sufficient statistic to know the individual's current preferences over redistribution. The currently poor may oppose redistribution if they expect their own income to improve in the future (Benabou and Ok, 2001, provide theoretical and empirical evidence that the POUM hypothesis works to limit the extent of redistribution in democracies). There is an obvious parallel between the POUM hypothesis and what we referred to as the signal effect of others' incomes in Section 13.2.1.

A number of pieces of empirical work have correspondingly underlined the importance of both current and future income. Along these lines, Ravallion and Lokshin (2000), using Russian microdata, were the first to show that self-assessed expected own social mobility, or the belief of being on a rising income trajectory, leads to lower demand for redistribution. Alesina and La Ferrara (2005) showed that preferences for redistribution are sensitive to the objectively measured future gains and losses that would result from it (again, this is consistent with pure self-interest). They also stressed the importance of mobility as an objective measure of the future expected gains and losses resulting from redistribution. In particular, there is a negative relationship between upward mobility (defined as the individual's own job prestige being higher than that of their father) and preferences for redistribution.[26] A subjective measure of whether the respondent says that he and his family "have a good chance of improving their standard of living" is very strongly negatively correlated with support for redistribution. Cojocaru (2014b) analyzed data from the second wave of the LiTS survey (the data are from 2010), and showed that preferences for redistribution are indeed linked to future upward mobility. The demand for redistribution is lower among those who are poor today but expect to be rich[27] in 4 years' time than for those who expect to be poor at both points in time. In line with the original POUM hypothesis, this finding only holds for those with lower levels of risk aversion (from a question on whether the individual would sell their car to buy insurance against a catastrophic drought).

[25] There is also a lively literature that emphasizes not necessarily whether I myself will benefit from redistribution, but also whether "people like me" are likely to do so. A recent survey of ethnic diversity and preferences for redistribution is provided by Stichnoth and Van der Straeten (2013).

[26] The survey in Alesina and Giuliano (2010) emphasizes the role of the past in general, both the individual's own past and the country's history.

[27] Poor and rich (now and future) are derived from the individual's response to a question about which decile of the income distribution they are at now and expect to be in four years' time. The poor (rich) are those who give an answer that is under (over) the average answer for the population.

Guillaud (2013) used 2006 ISSP data covering 33 countries to show that income and occupation are both important predictors of redistributive preferences. Equally, downward social mobility (having a lower position on the social scale now relative to 10 years ago) increases the demand for redistribution, whereas upward social mobility reduces it. There is some evidence that the downward mobility coefficient is larger in size than that on upward mobility, as if individuals were loss-averse with respect to status.

Clark and D'Angelo (2013) analyzed 18 waves of BHPS data. They showed that higher social status is associated with less favorable attitudes to redistribution and the public sector, as is commonly found. However, they also found that upward mobility (relative to one's parents) is associated with more left-wing attitudes, which are shown to be translated into actual reported voting behavior.

As noted in Section 13.3.1, Alesina et al. (2004) showed that the effect of inequality on happiness is larger in value in Europe than in the United States. The explanation proposed in Alesina et al. is in terms of greater perceived social mobility in the United States than in Europe.

Measures of the demand for redistribution have also been shown to be correlated with the individual's view of the fairness of the income distribution (Corneo and Grüner, 2002; Luttens and Valfort, 2012). The former test the importance of fairness in determining preferences for redistribution via the answer to the ISSP question "How important is hard work for getting ahead in life?" with responses "essential," "very important," "fairly important," "not very important," and "not important at all." They show that there is a self-interested component, in that those who state that they would personally benefit from lower inequality are indeed in favor of redistribution, whereas those with higher incomes are against distribution. The estimated coefficient on their fairness variable, "hard work is key" (defined as providing one of the first three responses given earlier), is shown to be negative and significant in a preferences for redistribution regression.

Luttens and Valfort (2012) appealed to data from the WVS and the ESS. They showed that both own income and the individual's perception of fairness determine redistribution preferences. It is of interest to note that individuals in the United States seem to more sensitive to fairness considerations in determining redistribution than are European respondents.

Tóth and Keller (2011) considered data from the 2009 Poverty and Social Exclusion module of the Eurobarometer. They calculated a Redistributive Preferences Index (RPI) using Principal Component Analysis of five questions on redistribution. The values of this index were then correlated with both individual and country-level variables. The latter include estimates of the distribution of income matched in from LIS data. They showed that the RPI is higher for those with lower material status, those who expect the situation to deteriorate over the next 12 months, and those who do not think that the poor are lazy. They also considered a number of percentile distribution measures (P95/P5, P95/P50, and P50/P5), as well as the Gini coefficient. All of the three percentile ratio measures attract positive significant estimated coefficients, so that the desire to redistribute rises with inequality. Inequality at the top and bottom of the distribution seems to play an equal role here. Yamamura (2012) also showed that the prefecture-level Gini coefficient is positively associated with redistributive preferences in 7 years of Japanese GSS data, although with a significant effect only for the richer.

One perhaps salient point here is that the questions used to establish preferences for redistribution are very different from one survey to another, which hampers the comparability of the existing results. Attitudes to income inequality are measured as follows in the BHPS: "People have different views about the way governments work. The government should place an upper limit on the amount of money that any one person can make." Answers to this question are on a 1–5 scale, where 1 represents complete disagreement and 5 complete agreement. This is not a question about redistribution in general, but about pulling the top of the distribution down. In the ISSP, respondents are asked "On the whole, do you think it should or should not be the government's responsibility to reduce income differences between the rich and the poor?" with answers on a 1–4 scale. The relevant question in the WVS asks individuals to indicate, on a 1–10 scale, which of the two extremes they most agree with: "People should take more responsibility to provide for themselves" versus "The government should take more responsibility to ensure that everyone is provided for." As Luttens and Valfort (2012) noted, this does mix up concerns for the income distribution with perceptions of government efficiency. Last, the question in the ESS is similar to that in the WVS, asking individuals to choose between "Government should decrease taxes a lot and spend much less on social benefits and services" and "Government should increase taxes a lot and spend much more on social benefits and services."

As well as individuals' stated preferences for redistribution, a recent paper has provided intriguing evidence that the actual observed progressivity of the tax system is positively correlated with average national well-being. Oishi et al. (2012) used data from 54 countries in the 2007 Gallup World Poll. Respondents here reported three different kinds of well-being measures: Cantril's ladder of the worst to best possible life, and positive and negative daily experiences. The country averages of these scores were correlated with the progressivity of the national tax system from Worldwide-Tax.com (calculated as the highest minus the lowest marginal tax rates, or the difference in the tax rates of those earning 67% and 167% of the country mean income). Tax progression is positively correlated with subjective well-being (see their Figure 1). This is not a simple revenue effect, as both the overall tax rate and government spending are significant in the well-being regressions.

13.4.4 Only Self-Interest?

The view of others in the comparative view of the reference group is arguably rather a depressing one. Other people are a negative externality in that $Y_i > Y_j$ brings relative satisfaction and $Y_i < Y_j$ relative deprivation for individual i. However certain others may be relevant for the individual, but not viewed in this comparative way. Rather, as intimated in the Introduction, there may well be a sentiment of extended sympathy toward some groups. In a parallel to the comparative reference group, the individuals toward whom one behaves altruistically will be chosen by the individual and may well exclude certain groups in society.

This leads us to the discussion of altruistic behavior, whereby transferring one's own money to others not only increases the recipient's well-being but also that of the donor. Although it is commonplace that generous people record higher well-being scores, showing causality from the former to the latter is more difficult.[28] Luckily there are a number of pieces of research that have suggested such a causal link.

One way of establishing causality is to use experiments. There has been something of a cottage industry in using randomized allocations or natural experiments to look at the relationship between own income and subjective well-being. Dunn et al. (2008) built on the observed positive correlation between prosocial spending and subjective well-being by considering a randomized experiment in which some individuals are forced to be generous. In particular, experimental participants first reported their happiness. They are then given an envelope with either \$5 or \$20 to spend that day. Half are told to spend the money on themselves, and the other half on someone else. Happiness recorded later that same evening showed a significant subjective well-being margin in favor of those who spent on others. Importantly, when surveyed regarding what they thought would make them happy, a separate sample of respondents thought that spending on themselves would make them happier than spending on others; as such, individuals are not necessarily aware of the happiness benefits of altruism *ex ante*.

Aknin et al. (2013) made the same point more broadly. They first reported a positive correlation between prosocial spending and happiness in 136 countries from the Gallup World Poll. They also appealed to experimental analysis. In Canada and Uganda, individuals asked to recall a past instance of prosocial spending reported higher happiness scores than did those who were asked to recall a past instance of personal spending; equally, in India individuals asked to recall a past instance of prosocial spending reported higher happiness scores than did those who were not asked to recall past spending. Last, along the same lines as Dunn et al. (2008), participants in Canada and South Africa who were randomly assigned to buy items for charity reported greater positive affect than those who were assigned to purchase the same items for personal use.

Boehm and Lyubomirsky (2009) showed that individuals in a treatment group who were told to perform three extra acts of kindness a day experienced a sustained rise in happiness compared to a control group.

[28] The experimental approach in Konow and Earley (2008) shows that those with (previously elicited) happiness scores are subsequently more generous in dictator games.

It is a small step from monetary donations to others to volunteering in general, and Carpenter and Myers (2010) showed that the two are indeed correlated. Meier and Stutzer (2008) analyzed survey data around the time of German reunification, which led to a sharp reduction in volunteering opportunities in East Germany. Meier and Stutzer showed that the drop in subjective well-being was larger for those who had previously volunteered than for those who had not: A natural conclusion is then that volunteering caused well-being.

A vibrant research area of interest in this respect covers charitable giving. Individuals may give to charity either because they care about the recipients of their largesse or because they derive some process utility from the act of giving that is independent of the use to which their gift is put (which is what Andreoni, 1989, calls "impure altruism").[29] Konow (2010) appeals to a series of carefully designed experiments to show that giving to others cannot be only explained by the "warm glow" of the process utility, and that the student subjects are systematically more generous toward charities than toward fellow students. This latter holds even when the charities are not known to the subjects, avoiding any role for familiarity. Konow suggested an overarching role for context-dependent norms in the determination of giving to others, which he identified as equity and need in his experiments. Useful relevant symposia on charitable donations can be found in the June 2011 special issue of the *Journal of Public Economics* on Charitable Giving and Fundraising and the forthcoming book edited by Fack and Landais (2014).

As noted by Clotfelter (2014), charitable giving is a more important phenomenon in the United States than in other G7 countries. However, it does remain unclear whether such giving is always redistributional, in the sense of being aimed at the less well-off. A first point is that some charitable donations, especially among the richer, go to the arts or education. Perhaps even more saliently, charitable donations in the United States are regressive in terms of the percentage of income donated (see http://philanthropy.com/article/Interactive-How-America-Gives/133709/).

Section 13.2.1 described a number of pieces of research in the vein of the comparative reference group, whereby higher incomes among relevant others were associated with lower levels of subjective well-being. This correlation is not always found to be negative, however. A variety of contributions have found that satisfaction and the income of close neighbors are actually positively correlated. This is the case in survey data in Canada (Barrington-Leigh and Helliwell, 2008), China (Kingdon and Knight, 2007), and Denmark (Clark et al., 2009a). Although the nonexperimental protocol here makes interpretation more difficult (there are any number of reasons why people might be happier with richer neighbors, including tunnel effects or the provision of local public goods), these findings are consistent with empathy with respect to close neighbors.

[29] Alternatively, charitable giving may be seen as a good that endows status on the benefactor, as in Frank (2004).

Kranton et al. (2013) also underlined that individuals can be altruistic toward some individuals, but comparative with respect to others. Individuals in their experiment make a series of choices regarding income allocations between two subjects. These subjects can be the individual, a member of her own group, or a member of another group. These groups are determined either by political persuasion, or as "minimal groups" depending on a preference over two nearly identical lines of poetry, landscape images, and abstract paintings. The authors found considerable heterogeneity in social preferences and showed that individuals are less generous (or even downright destructive, as in Zizzo and Oswald, 2001), toward individuals outside their group, even when groups are essentially randomly formed.

The theoretical implications of altruism in terms of redistribution are analyzed in Hochman and Rogers (1969). In this case, some redistribution can make everyone better off. Hochman and Rogers considered transfers only from richer to poorer and that do not change the income ranking. Transfers are costless; there is no leaky bucket. One of the central aims of their analysis is to establish how the amount transferred depends on the income gap between the rich and the poor. They distinguished two salient cases, which depend on the "transfer elasticity." When this elasticity is zero, the same fixed sum is always transferred; when it equals one, then the amount transferred is proportional to the income gap between the rich and the poor. A calibration suggests that actual U.S. income tax rates are more consistent with the elasticity being one than zero.

13.5. CONCLUSION

As so often in economics, asking about the relationship between income inequality and individual attitudes looked to be a pretty simple question, but turned out to be remarkably more delicate to answer.

The broad question addressed here is why individuals should care about the distribution of income in a society. The first useful distinction is whether they figure in the society in question or not. In the former case, income inequality will have implications for both their own income and their income relative to others; this is the comparative view of the income distribution. In the latter case, individuals can evaluate a distribution of income dispassionately, as it were, as this distribution will have no implications for either their own absolute income or their relative income; this is the normative view of the income distribution.

As a broad conclusion, there is now a variety of types of evidence that are consistent with individuals caring about their income position relative to others. To that extent, individuals do indeed have social preferences. It is worth underlining the unanimity that individuals dislike earning less than others. The "comparative" response to earning more than others remains open to debate. There may well be something of an asymmetry here, with the well-being advantage of earning more than others being smaller in absolute value than the well-being loss of earning less than others (a type of comparative loss aversion). However, the more extreme version of this aversion, with individuals actually disliking earning more than others remains unsettled. In general, the well-being effect of a rise in inequality under the comparative lens is ambiguous: Some people will become richer than those in their reference group, others will become poorer.

In contrast to these comparative findings, the happiness literature on the normative view of the income distribution has provided a wide scattershot of findings. One obvious difficulty in any approach based on survey subjective well-being data is effectively controlling for relative income when estimating the correlation between happiness and the income distribution. Very few analyses do so and therefore provide some kind of compound correlation, which includes both comparative and normative elements. The experimental approach here has a notable advantage in being able to distinguish the two.

Our reading of the many empirical analyses is that others' income most certainly does affect individual well-being, certainly in a comparative sense and very likely normatively too. At the same time, there are many qualifications to any broad-brush conclusion. First, the source of the income under consideration is key, with a consistent finding that individuals are less accepting of income gaps between individuals that are seen to be undeserving. Second, individuals can have separate views of different income distributions: It is quite possible to be altruistic with respect to one group, but comparative with respect to another. In this sense, it is not clear that there is only one "attitude" to inequality. Nor is it clear that such attitudes are fixed over time. For example, preferences for redistribution depend (in a self-interested way) on the individual's perceived position in the income distribution and on the degree of empathy toward others. Research in psychology has suggested that younger cohorts are more likely to rate themselves as above average (Konrath et al., 2011) and are less empathic (Twenge et al., 2012). What may have been unacceptable in the past in terms of the distribution of income may become anodyne in the future.

Research in this area has appealed to contributions from a variety of fields of research, both within economics and across the social sciences. It is striking how little these various fields communicate with each other. Any attempt to integrate at least some of the revealed preference, experiment, and happiness approaches would surely be welcome.

Individuals do have attitudes toward income inequality, whether these be stated, revealed, or measured physiologically or neurologically. To this extent, at least, man is a social animal. There is unlikely to be agreement any time soon about the "right" degree of inequality. This will be tied up with the societal extent of jealousy, altruism, fairness, and values. That many of these concepts are of such interest across the social sciences bodes well for Volume 3 of this Handbook.

ACKNOWLEDGMENTS

Thanks to Tony Atkinson and François Bourguignon for their careful reading of our previous versions, and to Dimitris Alexander Mavridis for proactive research assistance. We are grateful to Dimitris Ballas, Christian Bjornskov, Frederic Carlsson, Alexandru Cojocaru, Aditi Dimri, Lucio Esposito, Francesco Farina, Ernst Fehr, Bob Frank, Carol Graham, John Helliwell, Barbara Jancewicz, Olof Johansson-Stenman, James Konow, David Masclet, Agnar Sandmo, Claudia Senik, Tim Smeeding, Jean Twenge, Bernard van Praag, Ruut Veenhoven and Katsunori Yamada for help and thoughtful comments.

REFERENCES

Abbink, K., Masclet, D., Van Veelen, M., 2009. Reference Point Effects in Antisocial Preferences, CBESS, Discussion Paper 09-03.
Abel, A.B., 1990. Asset prices under habit formation and catching up with the Joneses. Am. Econ. Rev. 80, 38–42.
Abel, A.B., 2005. Optimal taxation when consumers have endogenous benchmark levels of consumption. Rev. Econ. Stud. 72, 21–42.
Abeler, J., Altmann, S., Kube, S., Wibral, M., 2010. Gift exchange and Workers' fairness concerns: when equality is unfair. J. Eur. Econ. Assoc. 8, 1299–1324.
Akerlof, G.A., Kranton, R.E., 2000. Economics and identity. Q. J. Econ. 115, 715–753.
Aknin, L., Barrington-Leigh, C., Dunn, E., Helliwell, J., Burns, J., Biswas-Diener, R., Kemeza, I., Nyende, P., Ashton-James, C., Norton, M., 2013. Prosocial spending and well-being: cross-cultural evidence for a psychological universal. J. Pers. Soc. Psychol. 104, 635–652.
Alesina, A., Giuliano, P., 2010. Preferences for redistribution. In: Benhabib, J., Jackson, M., Bisin, A. (Eds.), Handbook of Social Economics, vol. 1A. North Holland, Amsterdam, pp. 93–131.
Alesina, A., La Ferrara, E., 2005. Preferences for redistribution in the land of opportunities. J. Public Econ. 89, 897–931.
Alesina, A., Di Tella, R., MacCulloch, R., 2004. Inequality and happiness: are Europeans and Americans different? J. Public Econ. 88, 2009–2042.
Almås, I., Cappelen, A., Lind, J., Sørensen, E., Tungodden, B., 2011. Measuring unfair (in)equality. J. Public Econ. 95, 488–499.
Alpizar, F., Carlsson, F., Johansson-Stenman, O., 2005. How much do we care about absolute versus relative income and consumption? J. Econ. Behav. Org. 56, 405–421.
Amiel, Y., Cowell, F.A., 1992. Measurement of income inequality. Experimental test by questionnaire. J. Public Econ. 47, 3–26.
Amiel, Y., Cowell, F.A., 1999. Thinking About Inequality. Cambridge University Press, Cambridge.
Amiel, Y., Cowell, F.A., 2002. Attitudes towards risk and inequality: a questionnaire-experimental approach. In: Andersson, F., Holm, H.J. (Eds.), Experimental Economics: Financial Markets, Auctions, and Decision Making. Kluwer, Dewenter, pp. 85–115.
Amiel, Y., Creedy, J., Hurn, S., 1999. Measuring attitudes towards inequality. Scand. J. Econ. 101, 83–96.
Amiel, Y., Cowell, F.A., Gaertner, W., 2012. Distributional orderings: an approach with seven flavors. Theory Decis. 73, 381–399.
Andreoni, J., 1989. Giving with impure altruism: applications to charity and ricardian equivalence. J. Polit. Econ. 97, 1447–1458.
Aronsson, T., Johansson-Stenman, O., 2008. When the Joneses' consumption hurts: optimal public good provision and nonlinear income taxation. J. Public Econ. 92, 986–997.
Ball, R., 2001. Incomes, Inequality and Happiness: New Evidence, Haverford College, mimeo.
Barrington-Leigh, C., Helliwell, J., 2008. Empathy and Emulation: Life Satisfaction and the Urban Geography of Comparison Groups, NBER Working Paper 14593.
Becker, G.S., 1974. A theory of social interactions. J. Polit. Econ. 82, 1063–1093.
Beckman, S.R., Formby, J.P., Smith, W.J., 2004. Efficiency, equity and democracy: experimental evidence on Okun's leaky bucket. Res. Econ. Inequal. 11, 17–42.

Bell, L., Freeman, R., 2001. The incentive to work hard: explaining work hour differences in the US and Germany. Labour Econ. 8, 181–202.
Bellemare, C., Kröger, S., van Soest, A., 2008. Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities. Econometrica 76, 815–839.
Benabou, R., Ok, E., 2001. Social mobility and the demand for redistribution: the POUM hypothesis. Q. J. Econ. 116, 447–487.
Berg, M., Veenhoven, R., 2010. Income inequality and happiness in 119 nations. In: Greve, B. (Ed.), Social Policy and Happiness in Europe. Edward Elgar, Cheltenham, UK, pp. 174–194.
Bergh, A., 2008. A critical note on the theory of inequity aversion. J. Soc. Econ. 37, 1789–1796.
Biancotti, C., D'Alessio, G., 2008. Values, Inequality and Happiness. Banca d'Italia, Economic Working Paper No. 669.
Binmore, K., Shaked, A., 2010a. Experimental economics: where next? J. Econ. Behav. Org. 73, 87–100.
Binmore, K., Shaked, A., 2010b. Experimental economics: where next? Rejoinder. J. Econ. Behav. Org. 73, 120–121.
Bjørnskov, C., Dreher, A., Fischer, J., 2008. Cross-country determinants of life satisfaction: exploring different determinants across groups in society. Soc. Choice Welfare 30, 119–173.
Bjørnskov, C., Dreher, A., Fischer, J., Schnellenbach, J., Gehring, K., 2013. Inequality and happiness: when perceived social mobility and economic reality do not match. J. Econ. Behav. Org. 91, 75–92.
Blanchflower, D.G., Freeman, R.B., 1997. The attitudinal legacy of communist labor relations. Ind. Labor Relat. Rev. 50, 438–459.
Blanchflower, D.G., Oswald, A.J., 2003. Does Inequality Reduce Happiness? Evidence from the States of the USA from the 1970s to the 1990s, Dartmouth College, mimeo.
Bloch, F., Rao, V., Desai, S., 2004. Wedding celebrations as conspicuous consumption: signaling social status in rural India. J. Hum. Resour. 39, 675–695.
Boehm, J.K., Lyubomirsky, S., 2009. The promise of sustainable happiness. In: Lopez, S. (Ed.), Handbook of Positive Psychology. second ed. Oxford University Press, Oxford, pp. 667–677.
Bolton, G.E., 1991. A comparative model of bargaining: theory and evidence. Am. Econ. Rev. 81, 1096–1136.
Bolton, G.E., Ockenfels, A., 2000. ERC: a theory of equity, reciprocity, and competition. Am. Econ. Rev. 90, 166–193.
Boskin, M.J., Sheshinki, E., 1978. Optimal redistributive taxation when individual welfare depends upon relative income. Q. J. Econ. 92, 589–601.
Bossert, W., D'Ambrosio, C., 2007. Dynamic measures of individual deprivation. Soc. Choice Welfare 28, 77–88.
Bossert, W., D'Ambrosio, C., Peragine, V., 2007. Deprivation and social exclusion. Economica 74, 777–803.
Bowles, S., Park, Y., 2005. Emulation, inequality, and work hours: was Thorstein Veblen Right? Econ. J. 115, 397–412.
Boyce, C., Brown, G., Moore, S., 2010. Money and happiness: rank of income, not income, affects life satisfaction. Psychol. Sci. 21, 471–475.
Brodeur, A., Flèche, S., 2013. Where the Streets Have a Name: Income Comparisons in the US, IZA, Discussion Paper No. 7256.
Brooks, R., Scott, I., Maklakov, A., Kasumovic, M., Clark, A., Penton-Voak, I., 2011. National income inequality predicts women's preferences for masculinized faces better than health does. Proc. R. Soc. B Biol. Sci. 278, 810–812.
Brosnan, S., de Waal, F., 2003. Monkeys reject unequal pay. Nature 425, 297–299.
Brown, G., Gardner, J., Oswald, A.J., Qian, J., 2008. Does wage rank affect employees' wellbeing? Ind. Relat. 47, 355–389.
Brown, P., Bulte, E., Zhang, X., 2011. Positional spending and status seeking in rural China. J. Dev. Econ. 96, 139–149.
Burgoon, B., 2013. Inequality and anti-globalization backlash by political parties. Eur. Union Polit. 14, 408–435.
Camerer, C.F., 2003. Behavioural Game Theory. Princeton University Press, Princeton, NJ.

Camerer, C.F., Fehr, E., 2004. Measuring social norms and preferences using experimental games: a guide for social scientists. In: Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R. (Eds.), Foundations of Human Sociality. Oxford University Press, Oxford, pp. 55–95.
Camerer, C.F., Fehr, E., 2006. When does "economic man" dominate social behavior? Science 311, 47–52.
Campbell, J.Y., Cochrane, J.H., 1999. By force of habit: a consumption-based explanation of aggregate stock market behavior. J. Polit. Econ. 107, 205–251.
Cappelli, P., Sherer, P.D., 1988. Satisfaction, market wages and labor relations: an airline study. Ind. Relat. 27, 56–73.
Card, D., Mas, A., Moretti, E., Saez, E., 2012. Inequality at work: the effect of peer salaries on job satisfaction. Am. Econ. Rev. 102, 2981–3003.
Carlsson, F., Daruvala, D., Johansson-Stenman, O., 2005. Are people inequality-averse, or just risk-averse? Economica 72, 375–396.
Carlsson, F., Johansson-Stenman, O., Martinsson, P., 2007. Do you enjoy having more than others? Survey evidence of positional goods. Economica 74, 586–598.
Carpenter, J.P., Myers, C.K., 2010. Why volunteer? Evidence on the role of altruism, reputation, and incentives. J. Public Econ. 94, 911–920.
Chakravarty, S.R., 1997. Relative deprivation and satisfaction orderings. Keio Econ. Stud. 34, 17–31.
Chakravarty, S.R., 2009. Inequality, Polarization and Poverty. Advances in Distributional Analysis. Springer, New York, NY.
Charness, G., Rabin, M., 2002. Understanding social preferences with simple tests. Q. J. Econ. 117, 817–869.
Cherry, T.L., Frykblom, P., Shogren, J.F., 2002. Hardnose the dictator. Am. Econ. Rev. 92, 1218–1221.
Clark, A.E., 2003. Inequality-Aversion and Income Mobility: A Direct Test, DELTA, Discussion Paper 2003-11.
Clark, A.E., D'Angelo, E., 2013. Social Mobility, Wellbeing and Redistribution, PSE, mimeo.
Clark, A.E., Flèche, S., Senik, C., 2014. The great happiness moderation. In: Clark, A.E., Senik, C. (Eds.), Happiness and Economic Growth: Lessons from Developing Countries. Oxford University Press, Oxford, pp. 32–139.
Clark, A.E., Frijters, P., Shields, M., 2008. Relative income, happiness and utility: an explanation for the Easterlin paradox and other puzzles. J. Econ. Lit. 46, 95–144.
Clark, A.E., Kristensen, N., Westergård-Nielsen, N., 2009a. Economic satisfaction and income rank in small neighbourhoods. J. Eur. Econ. Assoc. 7, 519–527.
Clark, A.E., Kristensen, N., Westergård-Nielsen, N., 2009b. Job satisfaction and co-worker wages: status or signal? Econ. J. 119, 430–447.
Clark, A.E., Masclet, D., Villeval, M.-C., 2010. Effort and comparison income. Ind. Labor Relat. Rev. 63, 407–426.
Clark, A.E., Oswald, A.J., 1996. Satisfaction and comparison income. J. Public Econ. 61, 359–381.
Clark, A.E., Senik, C., 2010. Who compares to whom? The anatomy of income comparisons in Europe. Econ. J. 120, 573–594.
Clark, A.E., Senik, C., 2014. Income comparisons in Chinese villages. In: Clark, A.E., Senik, C. (Eds.), Happiness and Economic Growth: Lessons from Developing Countries. Oxford University Press, Oxford, pp. 216–233.
Clotfelter, C., 2014. Charitable giving and tax policy in the U.S. In: Fack, G., Landais, C. (Eds.), Charitable Giving and Tax Policy: A Historical and Comparative Perspective. Oxford University Press, Oxford, forthcoming.
Cohn, A., Fehr, E., Goette, L., 2014. Fair wages and effort: evidence from a field experiment. Manag. Sci, forthcoming.
Cojocaru, A., 2014a. Fairness and inequality tolerance: evidence from the Life in Transition survey. J. Comp. Econ. 42, 590–608.
Cojocaru, A., 2014b. Prospects of upward mobility and preferences for redistribution: evidence from the Life in Transition Survey. Eur. J. Polit. Econ. 34, 300–314.
Collard, D., 1975. Edgeworth's propositions on altruism. Q. J. Econ. 85, 355–360.
Corazzini, L., Esposito, L., Majorano, F., 2012. Reign in hell or serve in heaven? A cross-country journey into the relative vs. absolute perceptions of wellbeing. J. Econ. Behav. Org. 81, 715–730.

Corneo, G., Grüner, H.-P., 2002. Individual preferences for political redistribution. J. Public Econ. 83, 83–107.
Cowell, F., Cruces, G.A., 2004. Perceptions of inequality and risk. Res. Econ. Inequal. 12, 97–133.
Daly, M., Wilson, M., Vasdev, S., 2001. Income inequality and homicide rates in Canada and the United States. Can. J. Criminol. 43, 219–236.
D'Ambrosio, C., Frick, J.R., 2007. Income satisfaction and relative deprivation: an empirical link. Soc. Indic. Res. 81, 497–519.
D'Ambrosio, C., Frick, J.R., 2012. Individual well-being in a dynamic perspective. Economica 79, 284–302.
Dawes, C., Fowler, J.H., Johnson, T., McElreath, R., Smirnov, O., 2007. Egalitarian motives in humans. Nature 446, 794–796.
Dawes, C., Loewen, P., Schreiber, D., Simmons, A., Flagand, T., McElreath, R., Bokemper, S., Fowler, J., Paulus, M., 2012. Neural basis of egalitarian behaviour. Proc. Natl. Acad. Sci. U. S. A. 109, 6479–6483.
DeBruine, L., Jones, B., Crawford, J., Welling, L., Little, A., 2011. The health of a nation predicts their mate preferences: cross-cultural variation in women's preferences for masculinized male faces. Proc. R. Soc. B Biol. Sci. 277, 2405–2410.
de Vries, R., Gosling, S., Potter, J., 2011. Income inequality and personality: are less equal U.S. states less agreeable? Soc. Sci. Med. 72, 1978–1985.
Di Tella, R., MacCulloch, R., 2008. Gross national happiness as an answer to the Easterlin paradox? J. Dev. Econ. 86, 22–42.
Di Tella, R., MacCulloch, R.J., Oswald, A.J., 2001. Preferences over inflation and unemployment: evidence from surveys of happiness. Am. Econ. Rev. 91, 335–341.
Dohmen, T., Falk, A., Fliessbach, K., Sunde, U., Weber, B., 2011. Relative versus absolute income, joy of winning, and gender: brain imaging evidence. J. Public Econ. 95, 279–285.
Duesenberry, J.S., 1949. Income, Saving and the Theory of Consumer Behavior. Harvard University Press, Cambridge, MA.
Dunn, E., Aknin, L., Norton, M., 2008. Spending money on others promotes happiness. Science 319, 1687–1688.
Dupor, B., Liu, W.-F., 2003. Jealousy and equilibrium overconsumption. Am. Econ. Rev. 93, 423–428.
Easterlin, R.A., 1974. Does economic growth improve the human lot? In: David, P.A., Melvin, W.B. (Eds.), Nations and Households in Economic Growth. Stanford University Press, Palo Alto, pp. 89–125.
Easterlin, R., Angelescu-McVey, L., Switek, M., Sawangfa, O., Zweig, J., 2010. The happiness-income paradox revisited. Proc. Natl. Acad. Sci. 107, 22463–22468.
Ebert, U., 1988. Measurement of inequality: an attempt at unification and generalization. Soc. Choice Welfare 5, 147–169.
Ebert, U., Welsch, H., 2009. How do Europeans evaluate income distributions? An assessment based on happiness surveys. Rev. Income Wealth 55, 801–819.
Eckel, C., Gintis, H., 2010. Blaming the messenger: notes on the current state of experimental economics. J. Econ. Behav. Org. 73, 109–119.
Ehrenberg, R., Bognanno, M., 1990. Do tournaments have incentive effects? J. Polit. Econ. 98, 1307–1324.
Engel, C., 2011. Dictator games: a meta study. Exp. Econ. 14, 583–610.
Engelmann, D., Strobel, M., 2000. An Experimental Comparison of the Fairness Models by Bolton and Ockenfels and by Fehr and Schmidt, Econometric Society World Congress 2000 Contributed Papers 1229.
Engelmann, D., Strobel, M., 2004. Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. Am. Econ. Rev. 94, 857–869.
Engelmann, D., Strobel, M., 2007. Preferences over income distributions. Experimental evidence. Public Finance Rev. 35, 285–310.
Esteban, J.-M., Ray, D., 1994. On the measurement of polarization. Econometrica 62, 819–851.
Fack, G., Landais, C. (Eds.), 2014. Charitable Giving and Tax Policy: A Historical and Comparative Perspective. Oxford University Press, Oxford.
Falk, A., Kosse, F., Menrath, I., Verde, P., Siegrist, J., 2013. Unfair Pay and Health, University of Bonn, mimeo.

Farina, F., Grimalda, G., 2011. A Cross-Country Experimental Comparison of Preferences for Redistribution, Department of Economic Policy, Finance and Development (DEPFID) University of Siena Working Paper No. 2/2011.
Fehr, E., Gächter, S., 2000. Cooperation and punishment in public goods experiments. Am. Econ. Rev. 90, 980–994.
Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition and co-operation. Q. J. Econ. 114, 817–868.
Fehr, E., Schmidt, K.M., 2003. Theories of fairness and reciprocity: evidence and economic applications. In: Dewatripont, M., Hansen, L.P., Turnovsky, S.J. (Eds.), Advances in Economic Theory, Eighth World Congress of the Econometric Society, vol. I. Cambridge University Press, Cambridge, pp. 208–257.
Fehr, E., Schmidt, K.M., 2010. On inequity aversion. A reply to Binmore and Shaked. J. Econ. Behav. Org. 73, 101–108.
Ferrer-i-Carbonell, A., 2005. Income and well-being: an empirical analysis of the comparison income effect. J. Public Econ. 89, 997–1019.
Ferrer-i-Carbonell, A., Ramos, X., 2014. Inequality aversion and risk attitudes. J. Econ. Surv, forthcoming.
Fershtman, C., Weiss, Y., 1993. Social status, culture and economic performance. Econ. J. 103, 946–959.
Fleurbaey, M., 1995. Equality and responsibility. Eur. Econ. Rev. 39, 683–689.
Fließbach, K., Weber, B., Trautner, P., Dohmen, T., Sunde, U., Elger, C., Falk, A., 2007. Social comparison affects reward-related brain activity in the human ventral striatum. Science 318, 1305–1308.
Fließbach, K., Phillipps, C., Trautner, P., Schnabel, M., Elger, C., Falk, A., Weber, B., 2012. Neural responses to advantageous and disadvantageous inequity. Front. Hum. Neurosci. 6, 165.
Förster, M., Tóth, I., 2014. Cross-country evidence of the multiple causes of inequality in the OECD area. In: Atkinson, A., Bourguignon, F. (Eds.), Handbook of Income Distribution In: vol. 2B. (Chapter 19), Elsevier, Amsterdam, pp. 1713–1827.
Frank, R.H., 1985. Choosing the Right Pond: Human Behaviour and the Quest for Status. Oxford University Press, London.
Frank, R.H., 1999. Luxury Fever. The Free Press, New York, NY.
Frank, R.H., 2004. What Price the Moral High Ground? Princeton University Press, Princeton, NJ.
Frijters, P., Leigh, A., 2008. Materialism on the march: from conspicuous leisure to conspicuous consumption? J. Soc. Econ. 37, 1937–1945.
Gaertner, W., Namazie, C., 2003. Income inequality, risk, and the transfer principle: a questionnaire-experimental investigation. Math. Soc. Sci. 45, 229–245.
Galí, J., 1994. Keeping up with the joneses: consumption externalities, portfolio choice, and asset prices. J. Money Credit Bank. 26, 1–8.
Graham, C., Felton, A., 2006. Inequality and happiness: insights from Latin America. J. Econ. Inequ. 4, 107–122.
Grosfeld, I., Senik, C., 2010. The emerging aversion to inequality. Evidence from Poland 1992–2005. Econ. Trans. 18, 1–26.
Guillaud, E., 2013. Preferences for redistribution: an empirical analysis over 33 countries. J. Econ. Inequ. 11, 57–78.
Güth, W., Kliemt, H., Levati, M.V., 2009. (Over-)stylizing experimental findings and theorizing with sweeping generality. Ration. Mark. Morals. 0, 239–249.
Hagenaars, A.J., 1986. The Perception of Poverty. North-Holland, Amsterdam.
Hagerty, M.R., 2000. Social comparisons of income in one's community: evidence from national surveys of income and happiness. J. Pers. Soc. Psychol. 78, 764–771.
Heffetz, O., 2011. A test of conspicuous consumption: visibility and income elasticities. Rev. Econ. Stat. 93, 1101–1117.
Helliwell, J.F., 2003. How's life? Combining individual and national variables to explain subjective well-being. Econ. Model. 20, 331–360.
Herreiner, D.K., Puppe, C., 2010. Inequality aversion and efficiency with ordinal and cardinal social preferences—an experimental study. J. Econ. Behav. Org. 76, 238–253.

Hey, J.D., Lambert, P.J., 1980. Relative deprivation and the Gini coefficient: comment. Q. J. Econ. 95, 567–573.
Hirschman, A., 1973. The changing tolerance for income inequality in the course of economic development. Q. J. Econ. 87, 544–566.
Hochman, H.M., Rogers, J.D., 1969. Pareto optimal redistribution. Am. Econ. Rev. 59, 542–557.
Hoffman, E., McCabe, K., Shachat, K., Smith, V., 1994. Preferences, property rights, and anonymity in bargaining games. Games Econ. Behav. 7, 346–380.
Hyman, H.H., 1942. The psychology of status. Arch. Psychol. 269, 5–91.
Jancewicz, B., 2012. Measurement of Income Inequality Re-Examined. Constructing Experimental Tests by Questionnaire. University of Warsaw, mimeo.
Jancewicz, B., 2014. Perception of income inequality: a multidimensional scaling study. Res. Econ. Inequal, 22, forthcoming.
Jasso, G., 2007. Studying justice: measurement, estimation, and analysis of the actual reward and the just reward. In: Törnblom, K., Vermunt, R. (Eds.), Distributive and Procedural Justice: Research and Social Applications. Ashgate, London, pp. 225–253.
Jiang, S., Lu, M., Sato, H., 2012. Identity, inequality, and happiness: evidence from urban China. World Dev. 40, 1190–1200.
Johansson-Stenman, O., Carlsson, F., Daruvala, D., 2002. Measuring future grandparents' preferences for equality and relative standing. Econ. J. 112, 362–383.
Kelley, H.H., 1965. Two functions of reference groups. In: Proshansky, H., Siedemberg, B. (Eds.), Basic Studies in Social Psychology. Holt, Rinehart, and Winston, New York, pp. 210–214.
Kingdon, G., Knight, J., 2007. Community, comparisons and subjective well-being in a divided society. J. Econ. Behav. Org. 64, 69–90.
Knight, J., Song, L., Gunatilaka, R., 2009. Subjective well-being and its determinants in rural China. China Econ. Rev. 20, 635–649.
Konow, J., 2010. Mixed feelings: theories of and evidence on giving. J. Public Econ. 94, 279–297.
Konow, J., Earley, J., 2008. The hedonistic paradox: is homo economicus happier? J. Public Econ. 92, 1–33.
Konrath, S., O'Brien, E., Hsing, C., 2011. Changes in dispositional empathy in American college students over time: a meta-analysis. Pers. Soc. Psychol. Rev. 15, 180–198.
Kranton, R., Pease, M., Sanders, S., Huettel, S., 2013. Identity, Groups, and Social Preferences, Duke University, mimeo.
Krawczyk, M., 2010. A glimpse through the veil of ignorance: equality of opportunity and support for redistribution. J. Public Econ. 94, 131–141.
Kroll, Y., Davidovitz, L., 2003. Inequality aversion versus risk aversion. Economica 70, 19–29.
Kuhn, P., Kooreman, P., Soetevent, A., Kapteyn, A., 2011. The effects of lottery prizes on winners and their neighbors: evidence from the Dutch postcode lottery. Am. Econ. Rev. 101, 2226–2247.
Kuziemko, I., Buell, R., Reich, T., Norton, M., 2014. "Last-place aversion": evidence and redistributive implications. Q. J. Econ. 129, 105–149.
Layard, R., 2005. Happiness: Lessons from a New Science. Penguin, London.
Leibenstein, H., 1950. Bandwagon, Snob and Veblen effects in the theory of consumers' demand. Q. J. Econ. 64, 183–207.
Levitt, S., List, J., 2007. What do laboratory experiments measuring social preferences reveal about the real world? J. Econ. Perspect. 21, 153–174.
Ljungqvist, L., Uhlig, H., 2000. Tax policy and aggregate demand management under catching up with the joneses. Am. Econ. Rev. 90, 356–366.
Loewenstein, G., Thompson, L., Bazerman, M., 1989. Social utility and decision making in interpersonal contexts. J. Pers. Soc. Psychol. 57, 426–441.
Loughnan, S., Kuppens, P., Allik, J., Balazs, K., de Lemus, S., Dumont, K., Gargurevich, R., Hidegkuti, I., Leidner, B., Matos, L., Park, J., Realo, A., Shi, J., Eduardo Sojo, V., Tong, Y., Vaes, J., Verduyn, P., Yeung, V., Haslam, N., 2011. Economic inequality is linked to biased self-perception. Psychol. Sci. 22, 1254–1258.
Luechinger, S., 2009. Valuing air quality using the life satisfaction approach. Econ. J. 119, 482–515.
Luttens, R., Valfort, M.-A., 2012. Voting for redistribution under desert-sensitive altruism. Scand. J. Econ. 114, 881–907.
Luttmer, E., 2005. Neighbors as negatives: relative earnings and well-being. Q. J. Econ. 120, 963–1002.

Macours, K., 2011. Increasing inequality and civil conflict in Nepal. Oxford Econ. Pap. 63, 1–26.
Macunovich, D., 2011. A Note on Inequality Aversion Across Countries, Using Two New Measures, IZA Discussion Paper No. 5734.
Manski, C., 2000. Economic analysis of social interactions. J. Econ. Perspect. 14, 115–136.
Mas, A., 2006. Pay, reference points, and police performance. Q. J. Econ. 121, 783–821.
Masclet, D., Villeval, M.-C., 2008. Punishment, inequality, and welfare: a public good experiment. Soc. Choice Welfare 31, 475–502.
Meier, S., Stutzer, A., 2008. Is volunteering rewarding in itself? Economica 75, 39–59.
Merton, R.K., Kitt, A.S., 1950. Contributions to the theory of reference group behaviour. In: Merton, R. K., Lazersfeld, P.F. (Eds.), Continuities in Social Research. Studies in the Scope and Method of "The American Soldier". Free Press, New York, NY, pp. 42–53.
Morawetz, D., Atia, E., Bin-Nun, G., Felous, L., Gariplerden, Y., Harris, E., Soustiel, S., Tombros, G., Zarfaty, Y., 1977. Income distribution and self-rated happiness: some empirical evidence. Econ. J. 87, 511–522.
Moulton, B., 1990. An illustration of a pitfall in estimating the effects of aggregate variables on micro units. Rev. Econ. Stat. 72, 334–338.
Mujcic, R., Frijters, P., 2013. Economic choices and status: measuring preferences for income rank. Oxford Econ. Pap. 65, 47–73.
Ng, Y.-K., 1987. Relative-income effects and the appropriate level of public expenditure. Oxford Econ. Pap. 39, 293–300.
O'Connell, M., 2004. Fairly satisfied: economic equality, wealth and satisfaction. J. Econ. Psychol. 25, 297–305.
Oishi, S., Kesebir, S., Diener, E., 2011. Income inequality and happiness. Psychol. Sci. 22, 1095–1100.
Oishi, S., Schimmack, U., Diener, E., 2012. Progressive taxation and the subjective well-being of nations. Psychol. Sci. 23, 86–92.
Osberg, L., Smeeding, T., 2006. "Fair inequality"? Attitudes toward pay differentials: the United States in comparative perspective. Am. Sociol. Rev. 71, 450–473.
Oshio, T., Kobayashi, M., 2010. Income inequality, perceived happiness, and self-rated health: evidence from nationwide surveys in Japan. Soc. Sci. Med. 70, 1358–1366.
Persson, T., Tabellini, G., 1994. Is inequality harmful for growth? Am. Econ. Rev. 84, 600–621.
Piketty, T., 1995. Social mobility and redistributive politics. Q. J. Econ. 110, 551–584.
Pirttilä, J., Uusitalo, R., 2010. A 'leaky bucket' in the real world: estimating inequality aversion using survey data. Economica 77, 60–76.
Ravallion, M., Lokshin, M., 2000. Who wants to redistribute? The tunnel effect in 1990's Russia. J. Public Econ. 76, 87–104.
Rothstein, B., Uslaner, E., 2005. All for all: equality, corruption, and social trust. World Polit. 58, 41–72.
Rozer, J., Kraaykamp, G., 2013. Income inequality and subjective well-being: a cross-national study on the conditional effects of individual and national characteristics. Soc. Indic. Res. 113, 1009–1023.
Runciman, W.G., 1966. Relative Deprivation and Social Justice. Routledge, London.
Schneider, S.M., 2012. Income inequality and its consequences for life satisfaction: what role do social cognitions play? Soc. Indic. Res. 106, 419–438.
Schor, J., 1992. The Overworked American: The Unexpected Decline of Leisure. Basic Books, New York, NY.
Schwarze, J., Härpfer, M., 2007. Are people inequality-averse, and do they prefer redistribution by the state? Evidence from German longitudinal data on life satisfaction. J. Soc. Econ. 36, 233–249.
Senik, C., 2004. When information dominates comparison. Learning from Russian subjective panel data. J. Public Econ. 88, 2099–2123.
Senik, C., 2009. Direct evidence on income comparisons and their welfare effects. J. Econ. Behav. Org. 72, 408–424.
Shall, L.D., 1972. Interdependent utilities and Pareto optimality. Q. J. Econ. 86, 19–24.
Shibutani, T., 1955. Reference groups as perspectives. Am. J. Sociol. 60, 562–569.
Sloane, S., Baillargeon, R., Premack, D., 2012. Do infants have a sense of fairness? Psychol. Sci. 23, 158–160.
Smyth, R., Qian, X., 2008. Inequality and happiness in urban China. Econ. Bull. 4, 1–10.
Sobel, J., 2005. Interdependent preferences and reciprocity. J. Econ. Lit. 43, 392–436.
Solnick, S., Hemenway, D., 1998. Is more always better? A survey on positional. Concerns. J. Econ. Behav. Org. 37, 373–383.

Solnick, S., Hemenway, D., 2005. Are positional concerns stronger in some domains than in others? Am. Econ. Rev. 95, 147–151.
Steijn, S., Lancee, B., 2011. Does Income Inequality Negatively Affect General Trust? Examining Three Potential Problems with the Inequality-Trust Hypothesis. University of Amsterdam, GINI Discussion Paper No. 20.
Stevenson, B., Wolfers, J., 2008. Economic growth and subjective well-being: reassessing the Easterlin paradox. Brookings Pap. Econ. Act. Spring, 1–102.
Stichnoth, H., Van der Straeten, K., 2013. Ethnic diversity, public spending, and individual support for the welfare state: a review of the empirical literature. J. Econ. Surv. 27, 364–389.
Stigler, G.J., Becker, G.S., 1974. De Gustibus Non Est Disputandum. Am. Econ. Rev. 67, 76–90.
Takahashi, H., Kato, M., Matsuura, M., Mobbs, D., Suhara, T., Okubo, Y., 2009. When your gain is my pain and your pain is my gain: neural correlates of envy and schadenfreude. Science 323, 937–939.
Teyssier, S., 2012. Inequity and risk aversion in sequential public good games. Public Choice 151, 91–119.
Thaler, R.H., 1988. Anomalies: the ultimatum game. J. Econ. Perspect. 2, 195–206.
Tomes, N., 1986. Income distribution, happiness and satisfaction: a direct test of the interdependent preferences model. J. Econ. Psychol. 7, 425–446.
Tóth, I., Keller, T., 2011. Income Distributions, Inequality Perceptions and Redistributive Claims in European Societies, GINI Discussion Paper No. 7.
Tricomi, E., Rangel, A., Camerer, C., O'Doherty, J., 2010. Neural evidence for inequality-averse social preferences. Nature 463, 1089–1091.
Twenge, J., Campbell, W.K., Gentile, B., 2012. Generational increases in agentic self-evaluations among American college students, 1966–2009. Self Identity 11, 409–427.
Tyran, J.-R., Sausgruber, R., 2006. A little fairness may induce a lot of redistribution in democracy. Eur. Econ. Rev. 50, 469–485.
Van Praag, B.M.S., 1977. The perception of welfare inequality. Eur. Econ. Rev. 10, 189–207.
Van Praag, B.M.S., Baarsma, B.E., 2005. Using happiness surveys to value intangibles: the case of airport noise. Econ. J. 115, 224–246.
Van Veelen, M., 2012. Review of "A cooperative species: human reciprocity and its evolution" by Samuel Bowles and Herbert Gintis. J. Econ. Lit. 50, 797–803.
Vandendriessche, D., 2012. Inequality Aversion: Mediation of Needs and Social Comparisons. Masters thesis, Université de Paris 1.
Van de Stadt, H., Kapteyn, A., Van de Geer, S., 1985. The relativity of utility: evidence from panel data. Rev. Econ. Stat. 67, 179–187.
Van de Werfhorst, H.G., Salverda, W., 2012. Consequences of economic inequality: introduction to a special issue. Res. Soc. Stratif. Mobil. 30, 377–387.
Veblen, T., 1949. The Theory of the Leisure Class. George Allen and Unwin, London (Originally published 1899 by Macmillan New York).
Verme, P., 2011. Life satisfaction and income inequality. Rev. Income Wealth 57, 111–137.
Winkelmann, R., Winkelmann, L., 2010. Does inequality harm the middle-class? Evidence from Switzerland. Kyklos 63, 301–316.
Wu, Y., Zhang, D., Elieson, B., Zhou, X., 2012. Brain potentials in outcome evaluation: when social comparison takes effect. Int. J. Psychophysiol. 85, 145–152.
Yamada, K., Sato, M., 2013. Another avenue for anatomy of income comparisons: evidence from hypothetical choice experiments. J. Econ. Behav. Org. 89, 35–57.
Yamamura, E., 2012. Social capital, household income, and preferences for income redistribution. Eur. J. Polit. Econ. 28, 498–511.
Yitzhaki, S., 1979. Relative deprivation and the Gini coefficient. Q. J. Econ. 93, 321–324.
Zaki, J., Mitchell, J., 2011. Equitable decision making is associated with neural markers of intrinsic value. Proc. Natl. Acad. Sci. U. S. A. 108, 19761–19766.
Zizzo, D.J., Oswald, A.J., 2001. Are people willing to pay to reduce others' incomes? Ann. Econ. Stat. 63–64, 39–65.

CHAPTER 14

Inequality in Macroeconomics

Vincenzo Quadrini*, José-Víctor Ríos-Rull†
*University of Southern California, CEPR, Los Angeles, CA, USA
†University of Minnesota, Federal Reserve Bank of Minneapolis, CAERP, CEPR, NBER, Minneapolis, MN, USA

Contents

Handbook of Income Distribution, Volume 2B
ISSN 1574-0056, http://dx.doi.org/10.1016/B978-0-444-59429-7.00015-7

Abstract

We revise some of the main ways in which the study of aggregate performance of an economy overlaps with the study of inequality.

Keywords

Macromodels of inequality, Inequality dynamics, Inequality and financial markets, Political economy of inequality

JEL Classification Codes

E2, D31, B22

In a handbook devoted to income distribution, a chapter devoted to macroeconomics should start by clarifying the role of macroeconomics. Two of the main concerns about macroeconomics are aggregation and general equilibrium. The first ensures that the various sections of the economy aggregate, that is, by adding the incomes, wealth, and other variables of all households, we obtain the economywide value of these variables. The second macroeconomic concern is general equilibrium, that is, how changes in any section of the economy propagate to other sections of the economy via implied adjustments in prices and tax rates that are necessary to clear markets and balance the government's budget constraint. Although a large body of macroeconomic research abstracts from distributional considerations among individuals and households, a significant strand of studies are also concerned about the interaction between distribution and aggregate outcomes. In this chapter we will explore the possible interactions between distribution and the aggregate dynamics of the economy.

Since Bertola (2000) (the macroeconomics chapter in Volume 1 of this handbook series), many changes have taken place in the way macroeconomists deal with income inequality. One important change is that interest has shifted away from the study of the relation between inequality and long-term growth and has focused more on other aspects of macroeconomic performance. Perhaps the main reason for this change is that, in general, there is less concern about long-term growth. Now, the more popular view is that all advanced economies grow by about 2% annually. The main question is, what does it take for less-developed countries to accelerate the process of development and join the group of rich countries? As a result of these changes, macroeconomists have two main concerns with regard to inequality. One is what determines the joint distribution of earnings (or labor income) and wealth, and the other is how the explicit account of empirically sound inequality shapes the answers to the standard questions in macroeconomics. The models typically used feature a large number of agents that differ in earnings, wealth, and, in some cases, other characteristics. Consequently, we find it convenient to separate the two main branches of macroeconomic studies in this vein: the branch primarily interested in understanding the sources or causes of inequality and the branch concerned with the consequences of inequality for the aggregate performance of the economy. Such a distinction is not always applicable, yet it provides a natural organization of the literature. We will also find it occasionally convenient to separate studies that are primarily interested in economic growth from those focusing on business cycles.

The outline of the chapter is as follows. We start in Section 14.1 with some facts on the U.S. income and wealth distribution that are relevant from a macroeconomic point of view. We look at both cross-sectional evidence and the changes observed in the last few decades. Although some of these facts are analyzed in more detail in other chapters of this handbook (for example, Chapters 7–9), it will be useful to summarize them here as they are the reference for some of the theories we will review in this chapter.

After summarizing the main empirical facts about the income and wealth distribution, we take a look in Section 14.2 at how macroeconomists make sense of these facts. First, we show how the distribution of wealth is determined given an exogenous process for earnings. After reviewing the models used by macroeconomists to examine this question and their success in replicating the wealth distribution observed in the data, we turn to models of endogenous determination of earnings. We look at models of human capital investment to determine why some people are more successful than others, in the sense of earning higher labor incomes. Thus, we will look at earnings inequality not as a purely stochastic process (luck) but as an outcome of different mechanisms such as investment in human capital (for example, education) or the higher relative demand of certain skills (affecting the relative prices of certain skills compared with other skills). The section concludes with a look at how occupational choices can also determine labor earnings. This section is concerned with inequality as a permanent or steady-state phenomenon, and the occupational choice part is informed by the bad employment performance of the Great Recession, which includes business-cycle aspects.

Next we turn to the dynamics of inequality. Section 14.3 studies how inequality may change both over the business cycle and over a longer horizon. Here we consider a simple model in which factor shares—of capital and labor—can change.

Section 14.4 deals with what is possibly one of the most exciting ways in which macroeconomics and inequality interact: the role of financial markets or, more specifically, financial frictions. We start by looking at how the ability to borrow shapes the income and wealth distribution (and the allocation efficiency) by reallocating investment funds to entrepreneurs that are efficient and reliable, but not always both. We then turn to how wealth inequality is shaped by borrowing ability even when the rate of return of savings is equated across households. First, we look at how the sheer ability to borrow shapes inequality, and then we consider endogenous theories of borrowing where financial frictions arise from the institutional environment. We also look at various extensions of these ideas where the frictions are endogenous. In addition to exploring the effects of financial frictions on inequality, we look at the long-term effects on the performance of the economy, including some issues that have become of concern to macroeconomists, such as implications for global imbalances.

In Section 14.5 we analyze how the political system interacts with inequality to yield different policies that have an impact on the aggregate performance of the economy. People have different views about the desirability of alternative economic policies that depend on the position of individuals in the economywide distribution of income. The aggregation of individual preferences leads to the choice of particular policies. As the distribution of income changes, so does the choice of policies, which in turn affect the aggregate performance of the economy.

Section 14.6 concludes the chapter with a global assessment of what may be behind some of the changes observed in the last few decades.

Finally, a note of caution and a disclaimer. Throughout the chapter, we make use of various theoretical models that, for expositional purposes, are kept simple. Although this makes the intuitions of the basic mechanisms easy to understand, it also implies that these models may not be completely suited to address quantitative questions. Therefore, even if we often illustrate the properties of the model quantitatively, we should be careful in interpreting the simulation numbers as they are often intended to provide a qualitative, rather than quantitative, assessment of the model. The disclaimer is about the necessary incompleteness of this chapter. As much as we have tried to provide a general presentation of the studies that deal with inequality in macroeconomics, covering all possible subjects is impossible. There are many topics that we do not review. For example, we exclude studies that introduce behavioral elements in macroeconomic analysis. In part, this is motivated by our limited expertise in these subjects. We have also avoided for the most part the study of inequality in developing countries. We have marginally touched on issues such as the impact of the rise in inequality on the U.S. economy, the macroeconomic causes for the rising inequality, and globalization and inequality among others. Perhaps, more importantly, we have only scratched the surface of how income inequality translates into consumption inequality, which is what most economists think is what really matters. The topic of income inequality is quite vast, and different authors would write it quite differently; in fact, Thomas Piketty's chapter in this volume includes some macro modeling of the wealth distribution with a very different flavor than this chapter.

14.1. SOME FACTS ON THE INCOME AND WEALTH DISTRIBUTION

Here we outline some general features of the Lorenz curves for earnings and wealth and their correlation and persistence over a medium span of 5–10 years for both individuals and across generations. We draw data from Díaz-Giménez et al. (2011), Kuhn (2014), and Budría et al. (2001) for the United States.[1] The data come from the Survey of Consumer Finances (SCF). Although the facts are for the U.S. economy, they may apply in varying degrees to other countries. In general, the United States is a more extreme version of the other developed countries in the sense that it is characterized by higher inequality.

Table 14.1 shows the shares of earnings (the part of the income that can be attributed to labor), income (earnings plus capital income plus government transfers before taxes), and wealth (both financial and real assets, but not defined benefits pensions). See Díaz-Giménez et al. (2011) for details about the definition of these variables. As we can see, a large number of households have zero or negative earnings, almost two-thirds of all earnings come from the top quintile, and the top 1% receive almost 20% of all earnings. Our definition of earnings includes part of self-employment income that is imputed as labor income.[2] Because self-employment income can be negative, several households have zero or even negative earnings in our sample. Negative earnings contribute to a much higher Gini index compared to other measures provided in the literature, such as those that result from using earnings data from the Current Population Survey (CPS). However, correcting the CPS data by tax information could lead to Gini indexes that exceed 0.6 (see Alvaredo, 2011 and the discussion in Chapter 9). Wealth is more concentrated than income, and the poorest quintile holds negative wealth. Furthermore, more than 85% of the wealth is held by the richest quintile and more than one-third of all wealth by the richest 1%. Table 14.2 shows a few measures of dispersion that are useful to keep in mind.

The properties of the distribution of earnings, income, and wealth (total (Wea) and excluding housing (N–H–W)) have changed in the last few years. Tables 14.3–14.5 show the values of a few measures of concentration for 1998, 2007, and 2010. For earnings, the Ginis, the coefficient of variations, the various ratios involving the median and the shares of top groups have all increased, most of them monotonically. For income, the picture presented by the SCF is muddier. The Gini seems unchanged, with some measures indicating an increase in inequality and others a decrease. The same seems to have happened for wealth. While the Gini, the ratio of the 90th percentile to the median and of the

[1] In Krueger et al. (2010) various macroeconomists study inequality in a variety of other countries with a similar way of looking at the data as we do in this chapter.

[2] In fact, owing to the recession, the overall fraction of business income attributed to labor is larger in 2010 (93.4%) and in earlier waves of the SCF when it was around 85%. See Díaz-Giménez et al. (2011) for details on how to calculate such fraction.

Table 14.1 Distribution of earnings and net worth in the U.S. economy

	Bottom (%)			Quintiles					Top (%)			All
	0–1	1–5	5–10	1st	2nd	3rd	4th	5th	90–95	95–99	99–100	0–100
Shares of total sample sorted by earnings (%)												
Earnings	−0.1	0	0	−0.1	3.5	11	20.6	65.0	12.1	18.3	18.0	100
Income	0.8	0.4	0.9	6.5	8.5	10.5	17.8	56.7	10.3	16.5	15.9	100
Net worth	4.5	0.5	0.7	11.6	13	6.3	9.9	59.1	8.6	21.2	20.2	100
Shares of total sample sorted by net worth (%)												
Earnings	0.9	2.8	2.3	8.4	10.7	14.6	17.5	48.8	10.3	15.7	10.8	100.0
Income	0.8	2.6	2.2	8.4	10.0	13.9	17.6	50.1	10.2	15.3	12.2	100.0
Net worth	−0.3	−0.3	−0.1	−0.7	0.7	3.3	10.0	86.7	13.5	26.8	34.1	100.0

Data are from the 2010 Survey of Consumer Finances. Income includes all transfers including food stamps. Earnings are defined as the part of income earned by labor. Farm and business incomes are assigned split between labor (93.4%) and capital (6.6%) (a much lower capital share than in other years).
Source: Kuhn (2014).

Table 14.2 Concentration and skewness of the distributions for 2010

	Earnings	Income	Wealth
Coefficient of variation	3.26	3.45	6.35
Variance of the logs	1.41	0.92	4.65
Gini index	0.65	0.55	0.85
Top 1%/lowest 40%	210	67	47,534
Location of mean (%)	70	73	83
Mean/median	1.85	1.70	6.42

Source: Kuhn (2014).

Table 14.3 Changes in concentration

	Ginis				Coeff. of Variation			
	Ear	Inc	Wea	N–H–W	Ear	Inc	Wea	N–H–W
1998	0.61	0.55	0.80	0.86	2.86	3.56	6.47	7.93
2007	0.64	0.57	0.82	0.88	3.60	4.32	6.01	7.59
2010	0.65	0.55	0.85	0.89	3.26	3.45	6.35	7.70

Source: Kuhn (2014).

Table 14.4 Changes of relevant ratios involving the medians

	Median to 30th percentile				90th percentile to Median				Mean to Median			
	Ear	Inc	Wea	N–H–W	Ear	Inc	Wea	N–H–W	Ear	Inc	Wea	N–H–W
1998	2.80	1.71	4.00	4.54	3.18	2.87	6.88	12.56	1.57	1.62	3.95	7.66
2007	2.77	1.68	4.54	4.73	3.41	3.00	7.55	15.73	1.72	1.77	4.60	10.39
2010	3.30	1.64	5.24	4.11	3.79	3.10	12.37	23.33	1.85	1.71	6.42	13.18

Source: Kuhn (2014).

Table 14.5 Percentages of Total Earnings, Income, and Wealth of Selected Groups

	SCF Earnings			SCF Income			SCF Wealth		
	Top 10%	Top 1%	Top 0.1%	Top 10%	Top 1%	Top 0.1%	Top 10%	Top 1%	Top 0.1%
1998	43.5	16.1	1.7	42.8	17.4	6.1	68.6	33.9	12.5
2007	47.0	18.7	1.9	46.9	21.0	7.8	71.4	33.6	12.4
2010	48.4	18.0	1.7	44.5	17.2	5.6	74.4	34.1	12.3

Source: Kuhn (2014).

mean to the median have all gone up, the shares of the top 10%, 1% and 0.1%, have either remained stable or have gone down.

The modest evidence for an increase in inequality in income and wealth contrasts drastically with the picture reported by Piketty and Saez (2003), Piketty (2014), and Saez and Zucman (2014), who have used tax data. They have documented a big increase in inequality in the last few years in both income and wealth. Their evidence for income is direct as it comes straight from tax returns. The evidence for the evolution of wealth concentration in Saez and Zucman (2014) is indirect; it imputes the value of the assets that generate the reported capital income using the capitalization method with the rates of return for each class of assets that are obtained in the Flow of Funds. Yet it is quite persuasive. They suggest that the discrepancies that result between using the SCF and using tax data is due mostly to the top 0.1% of the top wealth holders. The SCF excludes the richest 400 households (the Forbes 400), and it is quite possible that even within income strata, response rates to the SCF voluntary questionnaire vary by income. The two sets of data are complementary, and the SCF is working hard to improve how it represents the very richest. There is likely to be a major update of the SCF that tries hard to improve in these dimensions. Hopefully, such improvements in the SCF will be available in the next few months, and we will get a much better picture of the characteristics of the very rich.

Turning to consumption inequality, there are some doubts about whether it has also increased. Using data from the Consumer Expenditure Survey, Krueger and Perri (2006) have documented that consumption inequality has increased only slightly. However, Attanasio et al. (2004) and Aguiar and Bils (2011) claim that the increase in consumption inequality has been more significant. One of the reasons these studies reach different conclusions is because they use different survey data. Krueger and Perri (2006) use data in which consumption comes from survey collecting interviews, and Attanasio et al. (2004) use diary data. In addition, Aguiar and Bils (2011) have argued that the observed reduction in the quality of the consumption data in terms of how much of aggregate consumption is recovered in the interviews is concentrated among goods that are mostly purchased by rich people and among the high income groups, and both features point to a larger increase in the underlying consumption inequality than what is obtained by using the data without special adjustments.

To summarize, over the last 10–20 years, the evidence points to a sizable increase in inequality.

14.2. MODELING THE SOURCES OF MACRO INEQUALITY

In this section we show what macroeconomics has to say about inequality. We start by exploring in Section 14.2.1 the implications of existing macro models for the distribution of wealth. Most of the models reviewed in this section start from the assumption that the process for earnings is exogenous: it can be stochastic, but it cannot be affected by individual decisions. In Section 14.2.1.3 we describe some of the theories in which the process for earnings is endogenous in the sense of being affected by individual choices. Because individual choices respond to policies, these models have interesting predictions about the impact of economic policies on the distribution of earnings.

14.2.1 Theories of Wealth Inequality Given the Process for Earnings

We start this section by emphasizing the limits of the neoclassical growth model with infinitely lived agents and complete markets in predicting wealth inequality. After reviewing the prediction of the overlapping generations model, we analyze models with incomplete markets. As we will see, the consideration of market incompleteness allows for more precise predictions about the distribution of wealth for given processes of individual earnings.

14.2.1.1 The Irrelevance of Income and Wealth Inequality in the Neoclassical Model

The deterministic neoclassical growth model says very little about income and wealth inequality. Note that we mean the neoclassical growth model in its modern meaning of incorporating fully optimizing saving behavior.[3] In an important article by Chatterjee (1994), reiterated later by Caselli and Ventura (2000), it is shown that any initial distribution of wealth is essentially self-perpetuating. To see this, consider the typical problem of a household $i \in \{1, \ldots, I\}$. Using recursive notation with primes denoting next period variables, the household's problem can be written as

$$v^i(a) = \max_{c,\, a'} u^i(c) + \beta_i v^i(a'), \tag{14.1}$$

$$\text{subject to} \quad c + a' = a(1+r) + \varepsilon_i w. \tag{14.2}$$

Here, $u^i(c)$ is a standard utility function (differentiable, strictly concave), $\beta_i \in [0, 1]$ is the discount factor, and ε_i is the household's endowment of efficient units of labor, which we assume constant for now. The necessary condition for optimality is

$$u_c^i(c_i) = \beta_i \ (1+r') \ u_c^i(c_i'), \tag{14.3}$$

where $u_c^i(c_i)$ is the marginal utility of consumption. In steady state, the allocation is constant over time, $c_i = c_i'$, and $r = r'$, which requires that the rate of return on savings is equal to the rate of time preference in every period, that is, $\beta_i = (1+r)^{-1}$. One implication is that, if households have interior first order conditions so that Equation (14.3) is satisfied with equality, then $\beta_i = \beta$ for all i. Otherwise, some households would reduce their assets as much as they can until they reach some lower bound that depends on the borrowing ability.

Because the rate of return in the neoclassical growth model is given by the marginal productivity of capital, we have that

[3] As such, the standard analysis of Stiglitz (1969) does not apply, since there saving behavior is postulated and not derived from first principles.

$$\beta^{-1} = 1 + r = F_K(K, N) - \delta, \tag{14.4}$$

where K is aggregate capital, $N = \sum_i \varepsilon_i$ is the aggregate effective labor (hours worked weighted by their efficiency), F is the production function, and δ is the constant rate of capital depreciation. In the neoclassical growth model, physical capital is the only form of wealth, so the following has to hold:

$$K = \sum_i a_i \tag{14.5}$$

where a_i are the assets held by household i. Note that these last three equations are the only ones imposed by the theory. It turns out that any distribution of wealth $\{a_i\}_{i=1}^{I}$ that satisfies Equations (14.4) and (14.5) is a steady state of this economy in which each individual household i consumes its income, $c_i = ar + \varepsilon_i w$. This is the sense in which the theory poses no constraints whatsoever on the distribution of a. Note that this is true no matter how efficiency units of labor (and hence earnings) are distributed across households. Nonseparability between consumption and leisure does not change this finding.

Small details qualify the behavior of the system outside a steady state. Under constant relative risk-aversion (CRRA) preferences, Equation (14.3) can be written as $\left(\frac{c_i'}{c_i}\right)^{\sigma} = \beta_i(1 + r')$, where $1/\sigma$ is the intertemporal elasticity of substitution. Depending on the joint distribution of earnings and wealth, the evolution of the wealth distribution is dictated by this equation and the budget constraint.

What other possibilities does the neoclassical growth model or its variants offer? Not many. Consider heterogeneity in the per period utility function. We have already noted that this does not change any steady-state consideration. Outside the steady state, the model just takes the initial wealth distribution and uses the first order conditions and the budget constraints to propagate the wealth distribution into the future, essentially dispersing or concentrating the wealth distribution without much endogenous action on the part of the model.

What about stochastic versions of these economies? With complete markets, all idiosyncratic uncertainty disappears (it is insured away), whereas the aggregate uncertainty is borne by those who are more willing to bear it. If such ability to bear the risk is increasing in wealth, then the model could generate some redistribution in response to aggregate shocks. But abstracting from aggregate uncertainty, we will see that the irrelevance result no longer applies when markets are incomplete (and agents continue to face idiosyncratic shocks). Before exploring the implications of incomplete markets, however, we briefly review the overlapping generations model.

14.2.1.2 Overlapping Generations Models and Wealth Inequality

In overlapping generations models, new households are born every period and live up to a certain number of periods J (they may also die earlier with some probability).[4] In what follows, we abstract from differences among households in any given age cohort and assume that the heterogeneity is only between cohorts. Households in age cohort j have earnings ε_j, which we take as exogenous. This specification can accommodate retirement and, with some extra work, government-provided Social Security (see Section 14.2.2 for theories of the determination of age-specific earnings). In a steady state, households solve the following problem:

$$\max_{\{c_j,\, a_{j+1}\}_{j=1}^{J}} \sum_{j=1}^{J} \beta_j u(c_j), \tag{14.6}$$

$$\text{subject to} \quad c_j + a_{j+1} = a_j(1+r) + w\varepsilon_j, \tag{14.7}$$

$$a_1 = 0, \tag{14.8}$$

$$a_{J+1} \geq 0. \tag{14.9}$$

Here, β_j is the specific weight that households place in the age-j utility. Note that households are born with no assets and cannot die with debts. Steady-state factor prices are r and w. The solution of the problem includes age-specific consumptions, c_j, and asset holdings, a_j, that satisfy the Euler equation

$$u_c(c_j) = \frac{\beta_{j+1}}{\beta_j}(1+r)u_c(c_{j+1}). \tag{14.10}$$

Steady-state factor prices are equal to the marginal productivities of a neoclassical production function with respect to aggregate capital, $K = \sum_{j=1}^{J} A_j$, and labor, $N = \sum_{j=1}^{J} \varepsilon_j$. We are using capital A_j to denote the assets of households of age j of which there are many, making it an aggregate variable which explains the use of capital letters.

When mapping these models to the data, we calibrate the earnings profile to have an inverse U shape as in the data (even after including Social Security payments). If we pose a constant discount rate, that is, we substitute β_j with β^j in Equation (14.6), as most researchers do, the model generates wealth holdings with an inverse U shape that typically peaks a little beyond 60 years of age. From that point on, the model predicts a slow but certain depletion of assets until death. Because in equilibrium household wealth has to add up to capital, households have to save during their finite lifetime to accumulate the whole capital stock. Although the prediction of the overlapping generations model in terms of lifetime wealth is broadly consistent with the data, the prediction for lifetime consumption is not. The strong incentive to save, together with the Euler equation,

[4] Sometimes the literature uses the term "overlapping generations model" for environments in which new agents are born every period and die with some probability at any point in the future. We refer to this particular environment as the Blanchard–Yaari model (Blanchard, 1985; Yaari, 1965), which is very similar mathematically to the infinitely lived model.

implies that $c_{j+1} > c_j$ for all j. In the data, however, consumption is also hump shaped. Various approaches are proposed in the literature to get around this shortcoming. They include demographic shifters, nonseparable leisure in the utility function (Auerbach and Kotlikoff, 1987; Ríos-Rull, 1996), existence of both durable goods and incomplete financial markets (Fernandez-Villaverde and Krueger, 2011), borrowing constraints and low rates of return (Gourinchas and Parker, 2002), and others.

With stochastic mortality, the model produces identical predictions as long as there is a market for annuities (which are available even if scarcely used). To see why, consider the probability of surviving between ages j and $j+1$, which we denote by φ_j. The survival probability multiplies the discount factor $\frac{\beta_{j+1}}{\beta_j}$, capturing the fact that the household gets utility only if alive.[5] Fairly priced (i.e., issued at zero expected cost) annuities imply that households save by purchasing them, and one unit of savings today yields $\frac{1}{\varphi_j}(1+r)$ units of the good tomorrow if the household survives and zero otherwise. Clearly, this asset dominates noncontingent investment of savings and the budget constraint (14.7) becomes

$$c_j + \frac{a_{j+1}}{\varphi_j} = a_j(1+r) + w. \tag{14.11}$$

It can be verified that with these modifications to discounting and the budget constraint, we obtain the same first order conditions as in Equation (14.10).

If we assume that there are no annuities, as in Hansen and Imrohoroglu (2008), we have to make some assumption about the allocation of the assets left by the deceased households. There are various options. One possibility is to assume that any household is like a pharaoh and assets are buried with their owners. The predictions of the model change a little relative to the basic model because there is now a smaller amount of total wealth due to the lower rate of return tilting the allocation toward young ages. Other options include the assumption that there is a 100% estate tax (with implications identical to that of the pharaoh model except for the use of public revenues) or that the assets of the deceased go to those in a certain age group. If the assets are distributed equally among the households of certain age groups, the wealth distribution will present a hike at the age at which households inherit, which is not a feature of the data. A more attractive alternative that has not been directly explored is to build direct links between a dead household and a randomly chosen younger household that inherits the assets. In this case, there will be limited within-cohort inequality that results from differences in the timing of the death and the wealth of the ancestors.

What about versions of overlapping generations economies with aggregate shocks? With aggregate shocks, even if there are markets for one-period-ahead state-contingent

[5] In this model, there is nothing that the household can do to affect survival, so the relative value of being alive or dead is irrelevant. With a CRRA utility function with more curvature than log, the utility is negative and, implicitly, our formulation seems to indicate that the household would rather be dead than alive.

assets, there could be incomplete insurance because households that are not alive cannot insure each other. The answers depend first on the size of the shocks. For (small) business-cycle type shocks, there are no great differences between the allocations implied by complete or incomplete markets. Ríos-Rull (1996) and Ríos-Rull (1994) find that the allocations are almost identical with and without typical business-cycle shocks. Larger and persistent shocks are a different matter. For example, Krueger and Kubler (2006) study the role of Social Security in reducing market incompleteness across generations and do not find large effects. Glover et al. (2011) study the redistributional implications of the (most recent) recession and find that the loss of output and consequent drop in the price of assets affect the old generations more than the young ones. The intuition for this result is that the recent crisis has been associated with large drops in asset prices, including housing, and old generations own more assets than the young.

If markets for the insurance of idiosyncratic risks are not present and households can save only by holding noncontingent assets, the situation changes dramatically and the model has very tight predictions. We will see this in the next section.

14.2.1.3 Stationary Theories of Earnings and Wealth Inequality

When households do not have access to insurance against shocks, the accumulation of riskless assets acts as a mechanism that allows households to smooth consumption—saving in good times when earnings are above the mean and dissaving in bad times. This means that in environments in which households are subject to uninsurable risks, those that have been lucky and have enjoyed good realizations of the shocks are wealthier than those that faced adverse realizations. This type of ex post inequality has been widely studied in models in which the risk was on endowments or earnings and agents could save only in the form of non-state-contingent assets. The basic theory was first developed in Bewley (1977), and the general equilibrium and quantitative properties were studied later by İmrohoroğlu (1989), Huggett (1993), and Aiyagari (1994). These ideas have important applications such as those in Carroll (1997) and Gourinchas and Parker (2002).

Successive studies have extended these models to improve the ability to generate greater wealth inequality. Among these approaches are the addition of special earning risks (Castañeda et al., 2003), entrepreneurial risks (Angeletos, 2007; Buera, 2009; Cagetti and De Nardi, 2006; Quadrini, 2000), endogenous accumulation of human capital (Terajima, 2006), and stochastic discounting (Krusell and Smith, 1998). Because in these models inequality is endogenous, the degree of wealth concentration can be affected by policies. This opened the way to studies that investigate the importance of taxation policies for wealth inequality. Examples are Díaz-Giménez and Pijoan-Mas (2011), Cagetti and De Nardi (2009), and Benhabib et al. (2011).

We start by reviewing how to pose the process for earnings (Section 14.2.1.3.1), and then we describe the main features of the Aiyagari model (Section 14.2.1.3.2).

14.2.1.3.1 Stochastic Representation of Earnings

A large body of literature tries to provide a parsimonious representation of the stochastic processes for wages or earnings. This literature uses panel data to estimate a univariate process for labor income or earnings, sometimes at the level of individual earners and sometimes at the household level (which is more in line with the data used in Tables 14.1 and 14.2).[6] (See, for example, Guvenen, 2009 or Guvenen and Kuruşçu, 2010; Guvenen and Kuruşçu, 2012.)

One important feature to take into account, as we will see below, is that the most common data sets do not include the very rich. The SCF is designed to provide a better picture of the rich but, unfortunately, it has no panel dimension and therefore cannot be used to separate individual effects from shocks and other interesting property that affects the most appropriate representation of earnings as a stochastic process. A comparison between the properties of the cross section in both data sets gives an idea of the differences in the sample. Recent work using either tax data (Atkinson et al., 2011; DeBacker et al., 2011) or Social Security data (Guvenen et al., 2012) looks very promising in terms of including both the very top earners and information about the persistence of their earnings.

14.2.1.3.2 The Aiyagari (1994) Model

Consider an economy populated by many, in fact a continuum, of infinitely lived agents that can be of finitely many types $i \in I$. They are subject to shocks that cannot be insured. Without loss of generality, we pose that there are finitely many possible realizations of the shock $m \in M$ and that it follows a Markov chain with (possibly type-specific) transition matrix $\Gamma^i_{m,m'}$. For compactness, we write Γ to denote a block diagonal matrix in which each block is $\Gamma^i_{m,m'}$. For the most part, the shock refers to the agents' endowment of efficiency units of labor, so we denote the shock by $s \in S = \{s^1, s^2, \ldots, s^M\}$.

Households do not care for leisure and assess consumption streams through a per period utility function $u^i(c)$ with intertemporal discount factor β_i. The utility function and the discounting may be type specific.

We start by considering the most primitive financial structure in which households have access to saving only in one-period noncontingent assets, and they cannot borrow. To map the model to a real economy and to consider its empirical implications, we build the model economy on top of a neoclassical growth model with exogenous labor supply. With a Cobb–Douglas production function, the prices of capital (rental rate of capital) and labor (wage) depend only on the capital-labor ratio. Because the aggregate labor supply is constant, prices depend only on aggregate capital K, and we can express them as $r(K)$ and $w(K)$.

[6] The process could also be for wages, but given the low variability of individual variance of hours worked for primary earners, wages and earnings have similar properties.

We consider only steady-state equilibria in which households face a constant interest rate r and a constant wage w per efficiency unit of labor. This approach is common in these types of studies because it greatly simplifies the computational burden. In fact, by focusing on steady states, when we solve the individual problem we can ignore the evolution of the aggregate states and we only need to keep track of the individual states (household's type i, asset position a, and realization of the idiosyncratic shock m). Of course, by doing so we have to exclude from the analysis changes that affect the whole economy, such as aggregate productivity shocks or structural changes. The consideration of aggregate and recurrent shocks represents a major computational complication (see, for example, Krusell and Smith, 1998). However, if we restrict ourselves to explore the implications of a one-time completely unexpected shock (a somewhat oxymoronic term) or structural change, then the computation remains tractable. For simplicity of exposition, we limit the analysis here to steady-state comparisons with the caveat that in the real economy, the distribution will take a long time to converge to a new steady state.

The household's problem can be written as

$$v^i(m, a; K) = \max_{c, a'} u^i(c) + \beta_i \sum_{m'} \Gamma^i_{m,m'} v^i(m', a'; K), \tag{14.12}$$

$$\text{s.t.} \quad c + a' = ws^m + a(1+r), \tag{14.13}$$

$$a' \geq 0, \tag{14.14}$$

where the superscript i denotes the household's type. Because this does not vary over time, we wrote it outside the arguments of the value function. The first order condition is given by

$$u^i_c(c) \geq \beta_i(1+r) \sum_{m'} \Gamma_{m,m'} u^i_c(c'), \quad \text{with equality if } a'_i > 0. \tag{14.15}$$

Standard results show that this problem is well behaved and the solution is given by a function $a_i'(m, a; K)$. Moreover, when $\beta_i^{-1} < (1+r)$, it is easy to show that for all i and m, there is a level of wealth $\overline{a}$ such that $a_i'(m, \overline{a}) < \overline{a}$. This means that there is a maximum level of wealth accumulated by an individual household. Thus, the set of possible asset holdings is the compact set $A = [0, \overline{a}]$.

To describe the economy, we could use a list of households with their types, shocks, and assets along with their names, but it is easier to use a measure x. This measure tells us how agents have certain characteristics in the space (i, m, a). Then, aggregate capital, which is just the sum of the assets of all households, can be written as

$$K = \int a \, dx. \tag{14.16}$$

The measure x gives us all the information that we need. For example, the total amount of efficiency units of labor or aggregate labor input is equal to

$$N = \int s^m dx, \tag{14.17}$$

and the variances of both wealth and earnings are

$$\sigma_k^2 = \int (a - K)^2 dx, \tag{14.18}$$

$$\sigma_N^2 = \int (s^m - N)^2 dx. \tag{14.19}$$

To calculate the Gini index for wealth, we need to compute the Lorenz curve and then calculate its integral. Note that any point of the Lorenez curve, for example, its value at 0.99 denoted by $\ell_{0.99}$, is one minus the share of wealth held by the richest 1%. To compute $\ell_{0.99}$, we start by finding the threshold of wealth that separates the richest 1% from the rest of households. Once we have found the threshold, we compute the wealth held by those households with wealth above the threshold relative to total wealth. The Gini index is simply twice the area between the Lorenz curve and the triangle below the diagonal between 0 and 1. (See Figure 14.6, for instance.) Other inequality statistics are also readily obtained from x, including those pertaining to the joint distribution of earnings and wealth and their intertemporal persistence.

The Aiyagari model has unique predictions about wealth and income inequality for any specification of the process of earnings. Therefore, the determination of the properties of the earning process becomes the central issue in the application of this model to the data. Should we think of people as being all ex ante equal in the sense that there is only one i type and they differ only in the realization of the shock? Or should we think of people consisting of different ex ante types? In either case, how do we determine which process to use? We now turn to this issue.

As we have seen in Section 14.1, the distribution of wealth in the United States is highly skewed, with about one-third of all the wealth in the hands of a mere 1% of households. How did those households become so wealthy? In order to become rich, households need both motive and opportunity. The reason for the opportunity is clear: at some point, the households had to have high enough earnings to be able to save and accumulate high levels of wealth. Motive is also important: why should households save rather than consume if they are impatient? If high earnings are not going to be around forever, then prudent households would want to save for the bad times that are likely to lie ahead. The issue is whether motives and opportunities are big enough to generate the wealth concentration observed in the United States.

To choose the actual parameterization of the earnings process for the model, we first need a Markovian process for earnings. If the focus is on the U.S. economy, one possibility is to specify a process estimated using the Panel Study of Income Dynamics (PSID). This is what Aiyagari (1994) did in his seminal paper, which relied on existing empirical studies such as Abowd and Card (1987) and Heaton and Lucas (1996).

The results are disappointing. The first row of Table 14.6 shows the shares of wealth of key groups in the U.S. data, and the second row displays those same shares as predicted by a model in which the earning process is calibrated using PSID data. The red line in Figure 14.4 shows the associated Lorenz curve. There is very little inequality compared with the inequality observed in the U.S. economy. The shares of wealth of the top 1% and the top quintile generated by the model are 4% and 27%, respectively, whereas in the United States these shares are 34% and 87%. The use of alternative estimates of the earning process, such as those provided by Storesletten et al. (2001), improves the performance of the model but only marginally.

The model fails to replicate the high concentration of wealth observed in the data for many possible reasons. One obvious explanation is that the model misses important pieces; for example, the model ignores life-cycle heterogeneity with all the demographic complications of actual lives. Or it ignores the permanent characteristics of people as well as education and human capital acquisition. It also ignores the fact that lives are affected by many other types of shocks such as health or unforeseen expenditures.

Castañeda et al. (2003) take a different approach and argue that the reason for the failure is the misrepresentation of the process of earnings. The PSID sample does not include very rich households. A comparison of its data with that of the SCF in which the emphasis is on wealthy people shows a large mismatch. The PSID does a better job at including people outside the top 10% of income earners and asset holders, but it is not appropriate to capture the dynamic properties of the incomes earned by the top of the distribution. Based on this observation, Castañeda et al. (2003) propose to ignore the PSID and focus instead on the specification of a process for earnings where the cross-sectional dispersion is similar to the SCF but its persistence is engineered so that it replicates the main features of the wealth inequality.

The third row of Table 14.6 displays the wealth distribution of an economy where the earning process has been calibrated following the above criteria. As we can see from the table, the model replicates quite well (by construction) the empirical data. Comparing the two processes for earnings (at least in the parsimonious representation used in Díaz et al., 2003) is very useful. The version of their economy designed to replicate the properties of the original Aiyagari economy has three values for earnings that are essentially symmetric, as are the persistence properties of the process: the earnings of agents in the middle and top thirds are $1.28\times$ and $1.63\times$, respectively, the earnings of those of the bottom third. Households in both, the top and the bottom thirds, have a one-third probability of moving out of their current situation by the next period. This society is very equal. Moreover, encountering bad luck, that is, being sent to the bottom third, is not really that bad. Clearly, because there are few motives to save money in this society, households soon stop doing so and consume all of their income.

The process that replicates the U.S. wealth distribution is extremely different from the process just described. The bottom of the society is now almost *half* of all households. Moreover, once at the bottom, less than 1% of these households move up each year.

Table 14.6 Concentration of wealth of various economies

Economy	1st	2nd	3rd	4th	5th	Top 10%	Top 5%	Top 1%	Gini
2010 U.S. Data	−0.7	0.7	3.3	10.0	86.7	74.4	60.9	34.1	0.85
PSID Ear–Pers	3.7	10.1	17.0	25.1	44.1	26.7	15.5	4.0	0.41
SCF Ear–Wea	0.0	0.0	0.3	4.5	95.2	78.3	53.2	14.7	0.87
Emp–Unemp	10.4	16.2	19.6	23.4	30.4	16.8	9.1	2.3	0.19
Stochastic β	1.7	6.5	12.5	21.1	59.3	40.3	25.3	6.7	0.56

Aiyagari (1994) used the PSID Ear–Pers calibration, Castañeda et al. (2003) used the SCF Ear–Wea calibration, and Krusell and Smith (1998) used both the Emp–Unemp and the Stochastic β calibrations.
U.S. data source Kuhn (2014).

Households that make up the middle class, almost one-half of the population, earn 5 × more than those at the bottom and have an equal chance (1%) of moving up or down. Only 6% of the households are in the top earnings group, and their earnings are huge: 47 × those of the poor and 9 × those of the middle class. More than 8% of these households will move down in a given year. Although these particular values are somewhat arbitrary, they give an accurate sense of how extreme both motives and opportunities have to be in order to induce the U.S. wealth disparities in a model in which agents differ only in the realization of a common earnings process.

Krusell and Smith (1998) pursue a very different approach to get a suitable wealth distribution. Instead of tracking the behavior of earnings, they pose a simple employment/unemployment process to generate earnings inequality, and they assume that the discount rate of individual agents is also stochastic. Therefore, in addition to the idiosyncratic shock to earnings, they consider a second idiosyncratic shock to the discount rate. The earnings process alone generates almost no inequality because the only way to get rich is through remaining employed but without earning more than other employed workers. In the extension with stochastic discounting, they assume that β can take three values, $\{0.9858, 0.9894, 0.9930\}$, with a symmetric distribution that satisfies the following properties: (i) the average duration of the extremes is 50 years (so it lasts the length of the adult life of a person); (ii) the transition from extreme to extreme requires a spell in the middle; and (iii) the (stationary) size of the middle group is 80%. Interestingly, the model with stochastic discounting yields disarmingly similar inequality indexes as in the data (see the last row of Table 14.6).

Various papers used life-cycle models with idiosyncratic risk to study wealth inequality. An important early contribution is Huggett (1996). De Nardi (2004), Cagetti and De Nardi (2006), and Cagetti and De Nardi (2009) study the role of bequest, estate taxation, and entrepreneurship in shaping the wealth distribution.

14.2.2 Theories of Earnings Inequality

So far we have described how macroeconomists think of the distribution of wealth given the distribution of earnings. But what about the distribution of earnings itself? Where is it coming from? In general, we can think of the differences in earnings as resulting from a combination of heterogeneity in (i) innate abilities or ex ante luck that persists for the lifetime of the agent; (ii) ex post luck due to the realization of shocks that are not under the control of the agent; (iii) effort or occupational choice; and (iv) investment in human capital. In the model considered in the previous section, the heterogeneity in earnings was only a consequence of innate heterogeneity (captured by the agents' type) and ex post luck (captured by the Markov process for skills). In this section we make the earnings endogenous by allowing for the optimal choice of effort and investment in human capital. An important consequence of endogenizing the earning process is that the distribution of earnings can be affected by several factors including financial market development (which, for example, facilitates access to the financing of investment in human capital) and taxation policies (which, for example, affect the marginal decision of effort and investment in human capital).

Next we briefly describe three aspects of models with endogenous earnings: models in which earnings are the result of explicit choices of either learning by doing or learning by not doing, including education (Section 14.2.2.1); models in which not all types of labor are perfect substitutes (Section 14.2.2.3); and models with occupational choices in which agents decide which occupation to take (Section 14.2.2.5).

14.2.2.1 Human Capital Investments

A common approach to endogenizing the process for earnings is the assumption that human capital is endogenous and depends on the individual investment chosen by agents. If the return from the investment is stochastic, then agents will be characterized ex post by different levels of human capital and, therefore, unequal earnings. An interesting feature of this setup is that it generates a positive relation between the aggregate performance of the economy and the degree of inequality. More specifically, higher investment in human capital leads to higher income or growth or both, but also to higher inequality because investment amplifies the impact of idiosyncratic shocks.

We illustrate this point with a simple model without taking a stand on the issue of what type of investment yields higher human capital. The investment can be either the result of time and hence forgone output or leisure, or the result of effort that generates a disutility, or the result of investment in goods. In this sense and at this level of abstraction, it accommodates both learning by direct investment in schooling or more general human capital investment such as that pioneered by Ben-Porath (1967). It also captures the key mechanisms formalized in more recent studies such as Guvenen et al. (2009), Manuelli and Seshadri (2010), and Huggett et al. (2011). For an extensive discussion of the life-cycle human capital model, see von Weizsäcker (1993).

Consider an economy with a continuum of risk-neutral workers, each characterized by human capital h. Production, which in this simple model corresponds to earnings, is equal to the worker's human capital h. Individual human capital can be enhanced with investment captured by the variable y. Investment is costly in terms of either utility or output (given risk neutrality, they are essentially the same). We assume that the cost takes the form $\frac{\alpha y^2 h}{2}$ and human capital evolves according to

$$h' = h(1 + y\varepsilon'),$$

where ε' is an i.i.d. random variable with $\mathbb{E}\varepsilon' = \bar{\varepsilon}$. To simplify notation, we denote by $g(y, \varepsilon') = 1 + y\varepsilon'$ the gross growth rate of human capital.

Because the outcome of the investment is stochastic, the model generates a complex distribution of human capital among workers. In the long run, the distribution will be degenerate because at the individual level h follows a random walk. To make the distribution stationary and keep the model simple, we assume that workers die with probability λ in each period and are replaced by the same mass of newborn workers. To allow for ex ante heterogeneity or innate abilities, we also assume that newborn agents are heterogeneous in initial human capital. In particular, there are I types of newborn agents indexed by $i \in \{1, ..., I\}$, each of size x_0^i and with initial human capital h_0^i. The initial distribution of newborn agents satisfies $\sum_i x_0^i = \lambda$.

Because of the linearity assumption, it will be convenient to normalize by h the optimization problem solved by a worker. We can then write the problem recursively as

$$\omega = \max_y \left\{ 1 - \frac{\alpha y^2}{2} + \beta(1-\lambda)\mathbb{E}[g(y, \varepsilon')\omega'] \right\}, \tag{14.20}$$

where ω is the expected lifetime utility normalized by human capital h. The non-normalized lifetime utility is ωh. Of course, the linearity of the accumulation function is crucial here. If the new human capital was a Cobb–Douglas function of old human capital, as in the Ben-Porath (1967) model, the analysis would be more complex analytically.

The first order condition gives

$$\alpha y = \beta(1-\lambda)\bar{\varepsilon}\omega', \tag{14.21}$$

where $\bar{\varepsilon}$ is the average value of the stochastic variable ε.

Because the first order condition is independent of h, the investment variable y is constant over time, which in turn implies that the normalized lifetime utility for the worker, ω, is constant. Therefore, y and ω can be determined by the two equations that define the value for the worker and the optimal investment, that is,

$$\omega = 1 - \frac{\alpha y^2}{2} + \beta(1-\lambda)(1 + y\bar{\varepsilon})\omega, \tag{14.22}$$

$$\alpha y = \beta(1-\lambda)\omega\bar{\varepsilon}. \tag{14.23}$$

Given the distribution of human capital for newborn workers x_0 and the investment variable y, we can determine the economywide distribution of human capital (equal to the distribution of earnings) and compute a cross-sectional index of inequality. We focus on the square of the coefficient of variation, that is,

$$\text{Inequality index} \equiv \frac{\text{Var}(h)}{\text{Ave}(h)^2},$$

which can be calculated exactly in a steady-state equilibrium.

Before we do this, note that the mass or measure of agents of age $j+1$ is given by $\sum_i x_j^i = \sum_i x_0^i (1-\lambda)^j$ and the average human capital is equal to

$$\text{Ave}(h) = \sum_i x_0^i \sum_{j=0}^{\infty} (1-\lambda)^j \mathbb{E}_j h_j^i. \tag{14.24}$$

The index j denotes the age of the worker and i the cohort of newly born agents with human capital h_0^i. The population size of newborn agents of type i is x_0^i, and the total mass of newborn agents is $\sum_i x_0^i = \lambda$. Because workers survive with probability $1-\lambda$, the fraction who is still alive after j periods is $(1-\lambda)^j$.

The cross-sectional variance of h is calculated using the formula

$$\text{Var}(h) = \sum_i x_0^i \sum_{j=0}^{\infty} (1-\lambda)^j \mathbb{E}_j \left[h_j^i - \text{Ave}(h) \right]^2, \tag{14.25}$$

which has an interpretation similar to the formula used to compute the average h. Of course, for the variance to be finite, we have to impose some parameter restrictions. In particular, we need to impose that the death probability λ is sufficiently large and the return on human capital accumulation $\mathbb{E}[\varepsilon']$ is not too big.

Using Equations (14.24) and (14.25), Appendix A shows that the average human capital and the inequality index take the forms

$$\text{Ave}(h) = \frac{\lambda \bar{h}_0}{1-(1-\lambda)\mathbb{E}[g(\gamma,\varepsilon)]}, \tag{14.26}$$

$$\text{Inequality index} = \left[\frac{\sum_i x_0^i \left(h_0^i\right)^2}{\bar{h}_0^2} \right] \frac{[1-(1-\lambda)\mathbb{E}g(\gamma,\varepsilon)]^2}{1-(1-\lambda)\mathbb{E}g(\gamma,\varepsilon)^2} - 1, \tag{14.27}$$

where $\bar{h}_0$ is the aggregate human capital of newborn agents.

We can see from Equation (14.26) that the average human capital and, therefore, aggregate output are strictly increasing in the investment variable γ. This is intuitive given the structure of the model.

As far as the inequality index is concerned, Equation (14.27) shows that this results from the product of two terms. The first term in parentheses captures the ex ante inequality, that is, the distribution of human capital at birth. If all agents are born with the same human capital, this term is 1. However, if the initial endowment is heterogeneous (heterogeneity in innate abilities), then this term is bigger than 1. The second term in parentheses captures the inequality generated by investment. It is easy to show that this term, and therefore, the inequality index, are strictly increasing in γ. Because the average value of h is also strictly increasing in γ, we have established that there is a positive relation between macroeconomic performance and inequality. The intuition for this dependence is simple. If $\gamma = 0$, human capital for all workers will be equal to h_0^i and the inequality index is fully determined by the ex ante heterogeneity. As γ becomes positive, inequality increases for two reasons. First, because the growth rate $g(\gamma, \varepsilon)$ is stochastic, human capital will differ *within* the same age-cohort of workers. Second, because each age-cohort experiences growth, the average human capital will differ *between* different age-cohorts.[7] Both mechanisms are amplified by the growth rate of human capital, which increases in the investment γ.

Using this model, we can analyze how changes that have an impact on the incentives to invest in human capital affect macroeconomic performance and inequality simultaneously. An example is a change in income taxes.

Suppose that the government taxes income at rate τ. The equilibrium conditions (14.22) and (14.23) become

$$\omega = 1 - \tau - \frac{\alpha\gamma^2}{2} + \beta(1-\lambda)(1+\gamma\bar{\varepsilon})\omega, \tag{14.28}$$

$$\alpha\gamma = \beta(1-\lambda)\omega\bar{\varepsilon}. \tag{14.29}$$

A bit of algebra shows that γ is strictly decreasing in τ. Effectively, the tax reduces the value of human capital, ω, which in turn must be associated with a reduction in γ (see Equation 14.29). Then, we can see from Equations (14.26) and (14.27) that higher taxes reduce inequality but also reduce the average human capital. This mechanism captures, in stylized form, the idea of Guvenen et al. (2009) used to explain cross-country wage inequality. They argue that higher taxation of labor accounts for the wage compression and lower productivity in Europe relative to the United States. Note also that the effects of higher labor taxation in the short run would differ from those in the long run in environments like this. In this particular model, taxation has no short-run disincentive effects (they would exist, however, if leisure were valued). Taxation does have long-run effects because agents would invest less in human capital. Empirical studies based only on short-run data would miss these effects.

14.2.2.2 Human Capital Investment Versus Learning by Doing

The stylized model considered in the previous section can easily be extended to include learning by doing. To do so, we can simply interpret the variable γ as the fraction of time spent investing (in human capital) and $1-\gamma$ the fraction of time spent producing. Output is produced according to the function $h(1-\alpha\gamma^2/2)$, which is strictly decreasing and

[7] This is in addition to the differences in initial human capital among the I types of newborn agents.

concave in the time spent investing. The equation determining the evolution of human capital becomes

$$h' = h(1 + \gamma\varepsilon') + \chi(1 - \gamma).$$

The first term captures the time spent investing, whereas the second results from learning by producing. The analysis conducted so far extends trivially to this case. In particular, the two Equations (14.22) and (14.23) become

$$\omega = 1 - \frac{\alpha\gamma^2}{2} + \beta(1 - \lambda)[(1 + \gamma\bar{\varepsilon}) + \chi(1 - \gamma)]\omega, \tag{14.30}$$

$$\alpha\gamma = \beta(1 - \lambda)\omega(\bar{\varepsilon} - \chi). \tag{14.31}$$

A further extension is to assume that the return from learning by doing is stochastic, that is, χ is a stochastic variable. Also, we could consider the special case in which the evolution of human capital is determined only by learning by doing. This case is obtained by setting $\varepsilon = 0$. These extensions do not change the basic properties of the model illustrated in the previous subsection, including the analysis of the short- and long-run effects of labor income taxation.

14.2.2.3 Prices of Skills

So far we have presented a model in which there is only one type of human capital or skills. Individuals have different levels of human capital and, therefore, earn different incomes. In reality, different types of skills are combined together with physical capital to produce goods and services. If those skills are not additive, they have a relative price that may be changing, implying that the distribution of income also depends on those relative prices, which in turn depend on the relative supplies and demands of the various skills.

To fix these ideas, suppose that there are three types of agents according to their skill types, H_1, H_2, and H_3. Production takes place through the technology

$$(H_1 + AH_2)^{\theta} H_3^{1-\theta}.$$

Assuming that markets are competitive, the prices of the three types of skills are equal to their marginal productivities, that is,

$$\begin{aligned} W_1 &= \theta(H_1 + AH_2)^{\theta-1} H_3^{1-\theta}, \\ W_2 &= \theta A(H_1 + AH_2)^{\theta-1} H_3^{1-\theta}, \\ W_3 &= (1 - \theta)(H_1 + AH_2)^{\theta} H_3^{-\theta}. \end{aligned}$$

In this example, the relative prices for the three types of skills depend on three factors: (i) the relative supplies of the skill types; (ii) the parameter A determining the productivity of H_2 relative to H_1; (iii) the parameter θ determining the relative productivity between the aggregation of H_1 and H_2 on one side and H_3 on the other. For example, an increase in the parameter A, keeping constant the relative supplies of the three skills, increases the productivity of H_2 and H_3 but reduces the marginal productivity of H_1. This changes the distribution of income between the three groups. The change in A could be the result of particular technological progress. As we will see in Section 14.3, a similar idea has been used by Krusell et al. (2000) to explain the increase in the skill premium observed in the United States since 1980.

14.2.2.4 Search and Inequality

Where does workers' luck come from? Some economists think it is from the arbitrariness of the process that matches workers to jobs. The idea is that some firms are better than others, and these firms end up paying more for essentially identical workers. The argument relies on two considerations. The first is that certain frictions make it difficult for firms to get a worker. The second is that wages depend on the characteristics of both workers and firms.

We can discuss these ideas with the help of the basic labor market model (see Pissarides, 1990) in which firms are created through the random matching of job vacancies and unemployed workers. Workers have linear utility $\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t c_t$. Risk neutrality implies that the interest rate is constant and equal to $r = 1/\beta - 1$.

A firm is created by paying a cost κ_0 that entails a draw of a productivity level z from the distribution $F(z)$. After the initial draw, z stays constant over time. Then the firm has to post a vacancy at cost κ_1. If matched with an unemployed worker, the firm produces output z starting in the next period until the match is separated, which happens exogenously with probability λ. The firm can use only one worker. The number of newly formed matches is determined by the function $M(v, u)$, where v is the number of vacancies and u is the number of unemployed workers. The probability that a vacancy is filled is $q = M(v, u)/v$, and the probability that an unemployed worker finds occupation is $p = M(v, u)/u$. The second ingredient of this model is that wages are determined through Nash bargaining, where we denote by η the bargaining power of workers. A worker attached to a firm with productivity z is paid the wage $w(z)$ and the firm earns $z - w(z)$.

The value of a firm that has a worker can be written recursively as

$$J^1(z) = \left\{ z - w(z) + \beta(1-\lambda) J^1(z) \right\}, \tag{14.32}$$

which implies that the value is $J^1(z) = \frac{z - w(z)}{1 - \beta(1-\lambda)}$. A newly created firm has value

$$J^0(z) = \left\{ -\kappa_1 + \beta(1-\lambda)\left[q J^1(z) + (1-q) J^0(z) \right] \right\}, \tag{14.33}$$

which can also be expressed as $J^0(z) = \dfrac{-\kappa_1 + \beta(1-\lambda)\dfrac{(z - w(z))q}{1 - \beta(1-\lambda)}}{1 - \beta(1-\lambda)(1-q)}$.

The value of a worker employed by a firm with productivity z is

$$W(z) = w(z) + \beta[(1-\lambda)W(z) + \lambda U], \tag{14.34}$$

where U is the value if the worker does not have a job. Such value is given by

$$U = \overline{u} + \beta\left\{p\int W(z)F(\mathrm{d}z) + (1-p)U\right\}, \tag{14.35}$$

where $\overline{u}$ is the flow utility for the unemployed worker.

To derive the bargaining problem, let's define the following functions:

$$\hat{J}(z,w) = z - w + \beta(1-\lambda)J^1(z), \tag{14.36}$$

$$\hat{W}(z,w) = w + \beta[(1-\lambda)W(z) + \lambda U]. \tag{14.37}$$

These functions are, respectively, the value of a firm and the value of an employed worker, given an arbitrary wage w paid in the current period and future wages determined by the function $w(z)$. The actual wage function $w(z)$ is the solution to the problem

$$\max_w \left[\hat{J}(z,w) - J^0(z)\right]^{1-\eta}\left[\hat{W}(z) - U\right]^{\eta}. \tag{14.38}$$

Notice that the terms inside the brackets describe, respectively, what the firm and the worker would lose if they do not reach an agreement and break the match. Parameter η captures the bargaining power of workers. To reach an equilibrium, a couple of additional conditions are needed. One is free entry of firms, that is, the expected value of creating a firm, $\int J^0(z)F(dz)$, equals its cost, κ_0. To get a steady-state equilibrium, total firm creation has to be sufficient to create enough vacancies to replace the jobs of workers that join unemployment from job separation.

It is easy to see that the wage is an increasing function of z. In this fashion, a theory of wage inequality can arise from the sheer luck of matching with a very productive firm, even though there is nothing inherently different between two workers in different firms. Hornstein et al. (2011) proposed a new method to assess the quantitative importance of the wage dispersion induced by search frictions and found that it is very small. In fact, the actual dispersion is $20\times$ larger than the dispersion generated by the type of search frictions described here. To understand this finding, think of an intermediate step between a worker being matched with a firm and before the actual bargaining process takes place. In this step, the worker could forecast what the wage will be and could potentially choose whether to take the job or keep searching. The minimum wage makes the worker indifferent between accepting the job and continuing searching. Such a wage can be compared with the average wage that workers get. Hornstein et al. (2011) found that for empirically sound values of the parameters, the difference between the minimum and the average wages generated by search frictions was tiny.

14.2.2.5 Occupational Choice and Earnings Inequality

Workers' choice of occupation has recently come to the fore as a source of income inequality. Income inequality may occur not only because workers accumulate different levels of human capital, but also because they work in different occupations. As evidenced by Kambourov and Manovskii (2009b), among others, human capital is largely occupation specific. We will discuss how occupation choices can directly affect the return to human capital and wage growth. In another vein, some workers' occupations may make them more sensitive to cyclical dynamics and unemployment. As in Wiczer (2013), the occupation specificity of human capital makes workers less flexible in response to specific shocks during the business cycle, and this generates inequality across occupations in terms of unemployment rates, unemployment duration, and earnings.

To see the pathways through which occupation choices may affect earnings inequality, consider a simple model with occupations indexed by $j=\{1,\ldots,J\}$. Human capital is only imperfectly transferable between occupations. Therefore, the human capital of a worker with current human capital h in occupation j that switches to occupation ℓ becomes $h'=\omega(h,j,\ell)$.

This is the basic framework with which Kambourov and Manovskii (2009a) connected occupational mobility to wage inequality. Workers who remain in the same occupation experience the same wage growth. Workers who switch occupations lose human capital, that is, experience negative growth in earnings. Let's normalize $h=1$ for experienced workers. When a new employee arrives, the human capital is $\omega_{j,\ell}=\omega(1,j,\ell)$, and it takes one period to become experienced. Let $g(j,\ell)$ be the probability that a worker switches from j to ℓ, and x_j is the measure of workers in occupation j. The variance of wages is

$$\mathrm{var}(w)=\sum_j x_j \sum_\ell g(j,\ell)\left(\omega_{j,\ell}-\mathbb{E}w\right)^2.$$

Clearly, without switching occupations, the variance would be zero. But, occupational switching is not infrequent and has, in fact, been rising concurrently with the recent rise in earnings inequality—the probability of switching occupations rose by 19% from the 1970s to the 1990s.[8] Kambourov and Manovskii (2009a) connect the former to the latter, posing a common cause for both. If occupation-specific shocks are on the rise, they will affect wage inequality through two channels. Directly, they will increase the dispersion in wages of those attached to an occupation, but shocks will also increase switching and create more wage inequality. Because these shocks are difficult to observe directly, Kambourov and Manovskii (2009a) use the occupational switching behavior of workers to inform the underlying process that would generate such behavior; as switching rose,

[8] This probability uses the 1970 Census occupations definitions at the three-digit level and PSID data.

the shocks must have also amplified. With this identification logic, occupational switching accounts for 30% of the overall rise in earnings inequality.

Unobserved, occupation-specific shocks are certainly not the only hypothesis for why workers move to different occupations. Several authors, e.g., Papageorgiou (2009) and Yamaguchi (2012), propose that workers' wages depend on occupation-specific match quality that is learned only through the course of the match. In these papers, earnings inequality is exacerbated by occupational mismatch, which slows the wage growth for some workers. On the other hand, aggregate factors such as business-cycle pressures and unemployment may also increase occupational switching. Indeed, unemployed workers are $3\times$ more likely than employed workers to switch occupations. In this vein, we introduce a simple model with unemployment and search as motives to switch occupations.

The critical object to be determined in this environment is the stochastic finding rate, $\{p_{j,\ell}\}$, at which workers with occupation j find a job in occupation ℓ. Denote by $P=\{p_{j,\ell}\}_{j,\ell\in\{1,\ldots,J\}}$ the collection of all such finding rates. Denote by $g(j,\ell)$ the fraction of time spent searching for occupation ℓ by a worker with previous occupation j. We assume that looking for jobs in a particular area has decreasing returns at rate $\varphi<1$, so that not all workers from occupation j shift to the same type of new occupation ℓ. This simplification could be a stand-in for any number of more realistic elements of a model such as heterogeneous preferences. Clearly, the allocation of searching time determines the realized finding rate.

The characteristics of the equilibrium are going to depend on the type of wage-setting rule we use, but here we consider the simplest case in which workers earn their marginal product. Hence, the wage of a worker who has just switched occupations is $\omega_{j,\ell}$ and 1 for an experienced worker.

Let W be the value function for an employed worker and U for an unemployed worker. These functions are defined recursively as

$$\begin{aligned} W(\ell,P) &= \omega_{\ell,\ell}+\beta(1-\lambda)\mathbb{E}\,W(\ell,P')+\beta\lambda\mathbb{E}\,U(\ell,P'),\\ U(j,P) &= \max_{g(j,\ell)}\sum_{\ell=1,\ldots,J} g(j,\ell)^{\varphi}p_{j,\ell}\big(\omega_{j,\ell}+\beta\mathbb{E}W(\ell,P')\big)+\left(1-\sum_{\ell=1,\ldots,J} g(j,\ell)^{\varphi}p_{j,\ell}\right)\\ &\quad\times(\overline{u}+\beta\mathbb{E}U(j,P')). \end{aligned}$$

In this case, $g(j,\ell)$ is chosen so that, given wages and finding rates, the marginal return to search time satisfies

$$\begin{aligned} &g(j,j)^{\varphi-1}p_{j,j}\{1-\overline{u}+\beta\mathbb{E}[W(j,P')-U(j,P')]\}\\ &\quad=g(j,\ell)^{\varphi-1}p_{j,\ell}\big\{\omega_{j,\ell}-\overline{u}+\beta\mathbb{E}[W(\ell,P')-U(j,P')]\big\},\quad\forall\ell. \end{aligned}$$

To get a taste of the dynamics, suppose that $p_{j,j}$ falls. The indifference condition holds that search time toward this occupation falls so that $g(j,j)^{\varphi-1}$ will rise. This increases earnings inequality through two channels: (1) the increase in the unemployment rate among type j workers and (2) more of the new matches go to different occupations, where they produce only $\omega_{j,\ell} < 1$.

The model we have presented has the essential elements of that explored in Wiczer (2013) and, in tying it to data, he shows that recessions often bring a correlated change in P that hurts some occupations much more than others. In such a case, workers from the affected occupations can be unemployed for very long durations and keep the level of unemployment high for a long time. To allow for this result, Wiczer (2013) builds upon Kambourov and Manovskii (2009a) by introducing search frictions, so that the job finding rate is endogenous and affected by business-cycle conditions. As in the case of a typical matching function, the finding rate is reduced by *congestion*.

To extend our framework to endogenous matching frictions, following Wiczer (2013) let $p_{j,\ell} = p(z_\ell, g(j,\ell))$, where z_ℓ is a shock affecting hiring in occupation ℓ and the more workers looking for the same types of job lowers the finding rate, $\frac{\partial p}{\partial g} < 0$. Then, when $p_{j,j}$ falls, the probability of successfully switching occupations also falls. Whereas Kambourov and Manovskii (2009a) find their shocks to reconcile switching behavior, Wiczer (2013) maps his shocks to measured value added by occupation. Looking directly at productivity allows Wiczer (2013) to address business cycles in which unemployment and earnings dispersion across occupations increases even though search frictions prevent workers from mass switches into new jobs.

How to identify "occupations" in the data is still an open question. Whereas Wiczer (2013) uses two-digit occupation codes, Carrillo-Tudela and Visschers (2013) take a similar model but with a finer definition of occupation that highlights the interaction between occupation-specific skills and other job characteristics such as location. Hence, the position of a machinist in Detroit may be even more volatile than machinists in general. Both papers generate significant volatility in unemployment and earnings over the business cycle beyond search models that abstract from occupational heterogeneity.

14.3. THE DYNAMICS OF INEQUALITY

So far we have looked at how to build theories of inequality in earnings and wealth that aggregate into a macro model. Now we turn to the analysis of factors that affect the dynamics of inequality. We first consider in Section 14.3.1 changes in inequality that take place over the business cycle, and in Section 14.3.2 we analyze the dynamics over a longer horizon.

14.3.1 Inequality and the Business Cycle

A well-established feature of the business cycle is that the labor share of income is highly countercyclical. As shown in Figure 14.1, the labor share in the U.S. economy tends to increase during recessions. The figure also shows a declining trend in the labor share since

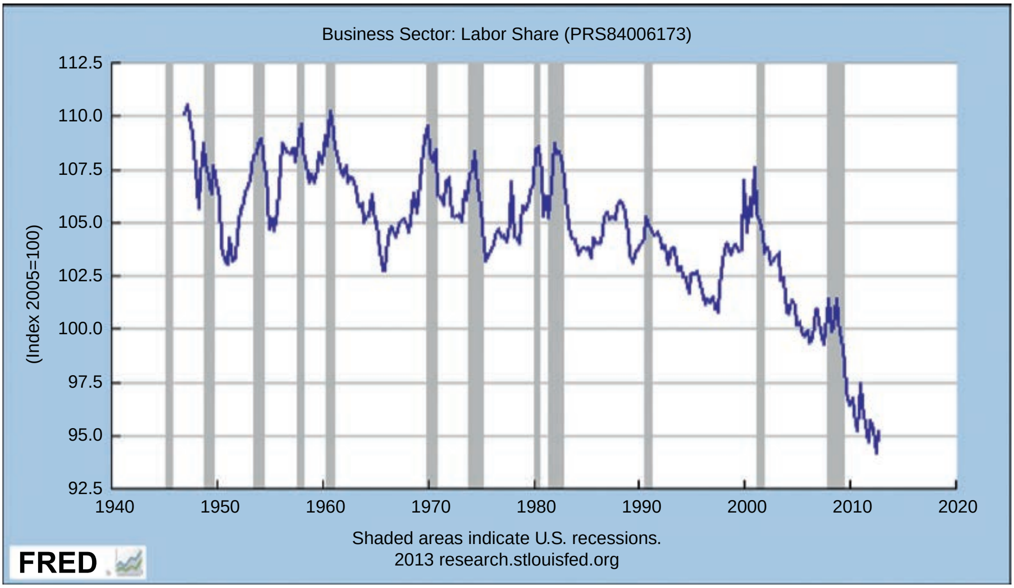


Figure 14.1 Labor share in the U.S. business sector as defined by the Bureau of Labor Statistics. *Source: U.S. Department of Labor: Bureau of Labor Statistics.*

the early 1980s. To the extent that agents are heterogeneous in the sources of their income—that is, some agents earn primarily capital incomes whereas others earn primarily labor income—there is significant redistribution over the business cycle.

To capture the cyclical properties of the labor share, we have to deviate from the standard neoclassical model with a Cobb–Douglas production function, because in this model the labor share is constant. In this section, we review some models in which the compensation of workers is determined through bargaining between employers and workers. Because the bargaining strength of workers depends on macroeconomic conditions, this mechanism has the potential to generate a labor income share that changes over the business cycle.

We start in Section 14.3.1.1 by using the search and matching model developed in Section 14.2.2.4 modified to allow for the study of the determination of the labor share. The modification associates the creation of firms with actual investors, which gives an explicit separation of labor and capital income. We look at two versions of this model: a simple one and another in which investors use both debt and external financing with bonds. An important property of this model is that shocks that affect the bargaining position of workers also affect the distribution of income as well as the macroeconomic impact of these shocks. We will consider two types of shocks: standard productivity shocks and shocks that affect access to credit. Then in Section 14.3.1.2, we will review the financial accelerator model in which the distribution is also interconnected with the business cycle. We will conclude this section by discussing the ability of these models to replicate the empirical properties of the data beyond the contemporaneous correlation.

14.3.1.1 The Determination of Factor Shares: Productivity Shocks, Bargaining Power Shocks, and Financial Shocks

Consider a version of the search and matching model described in Section 14.2.2.4 (Pissarides, 1987) in which the owners of firms—investors—are distinct from workers, but in which productivity is stochastic and common to all firms. Therefore, z is the same across firms and changes stochastically over time. We will focus on the distribution of income between investors and workers. Both types of agents have the same utility $\mathbb{E}_0\sum_{t=0}^{\infty}\beta^t c_t$.

As before, a firm is created when a posted vacancy is filled by an unemployed worker. A new firm produces output until the match is destroyed exogenously, which happens with probability λ, but now the level of output varies over time. The number of matches is determined by the matching function $M(v,u)$, where v is vacancies and u is unemployed workers. The probability that a vacancy is filled is $q=M(v,u)/v$, and the probability that an unemployed worker finds a job is $p=M(v,u)/u$. Wages are determined through Nash bargaining, where we denote by η the bargaining power of workers. We also consider the possibility that the bargaining power η may be stochastic. With these stochastic terms we rewrite the value functions for investors as

$$J^1(z,\eta)=z-w(z,\eta)+\beta(1-\lambda)\mathbb{E}J^1(z',\eta'|z,\eta), \tag{14.39}$$

$$J^0(z,\eta)=-\kappa_1+\beta(1-\lambda)\left[q\mathbb{E}J^1(z',\eta'|z,\eta)+(1-q)\mathbb{E}J^0(z',\eta'|z,\eta)\right], \tag{14.40}$$

and for workers

$$W(z,\eta)=w(z,\eta)+\beta\mathbb{E}[(1-\lambda)W(z',\eta')+\lambda U(z',\eta')], \tag{14.41}$$

$$U(z,\eta)=\overline{u}+\beta\mathbb{E}[pW(z',\eta')+(1-p)U(z',\eta')]. \tag{14.42}$$

After some rearrangement, the values for the firm and the worker can be written as

$$J^1(z,\eta)-J^0(z,\eta)=(1-\eta)S(z,\eta), \tag{14.43}$$

$$W(z,\eta)-U(z,\eta)=\eta S(z,\eta), \tag{14.44}$$

where $S(z,\eta)=J^1(z,\eta)-J^0(z,\eta)+W(z,\eta)-U(z,\eta)$ is the bargaining surplus that is split between the contractual parties, proportional to their relative bargaining power. The surplus function can be written recursively as

$$S(z,\eta)=z-\overline{u}+(1-\lambda)\beta\mathbb{E}S(z',\eta')-\eta\beta p\mathbb{E}S(z'\eta'). \tag{14.45}$$

Using the free-entry condition $\kappa_1=q\beta\mathbb{E}S(z',\eta')$, the sharing rules (14.43) and (14.44), and the functions (14.39), (14.41), and (14.42), we can derive the following expression for the wage:

$$w(z,\eta) = (1-\eta)\overline{u} + \eta z + \frac{\eta p \kappa_1}{q}.$$

14.3.1.1.1 Shocks to Productivity

Figure 14.2 plots the impulse responses of employment and the investor's share of income to a positive productivity shock under the heading baseline model. An economic boom is characterized by a larger share of income going to investors. However, the quantitative effects in terms of income distribution and employment are not large. The weak employment response is a well-known property of the matching model (see Costain and Reiter, 2008 or Shimer, 2005). What is interesting is that the inability of the model to generate large employment fluctuations is related to the inability of the model to generate large movements in the distribution of income. Because wages respond too quickly to

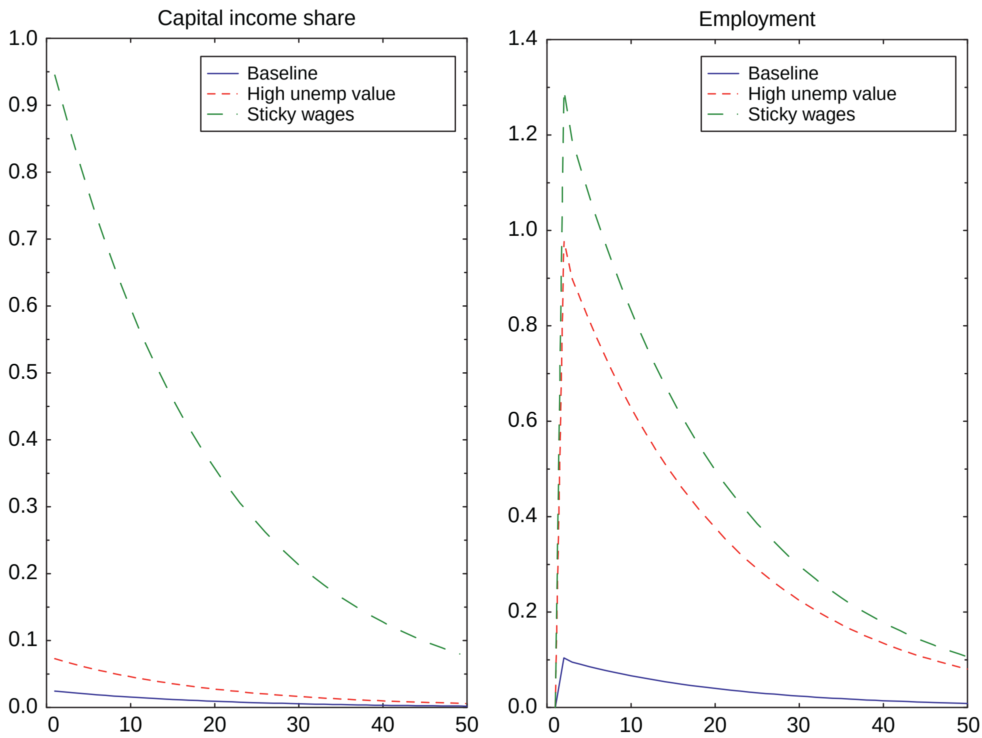


Figure 14.2 Impulse response to productivity shock. The common parameters to all versions of the model are $\beta = 0.985$, $\alpha = 0.5$, $\mathbb{E}z = 1$, $\rho_z = 0.95$, $\sigma_z = 0.01$. The remaining parameters $\overline{u}$, κ, λ, and A are chosen to achieve the following steady-state targets: a replacement rate of unemployment of 50% (95% in the model with high unemployment value), 10% unemployment rate, 93% probability of filling a vacancy, 70% probability of finding a job. The resulting values are $\overline{u} = 0.473$ (0.944 in the model with high unemployment value), $\kappa = 0.316$ (0.034 in the model with high unemployment value), $\lambda = 0.103$, and $A = 0.807$.

productivity, the share of income going to investors increases only slightly. As a result, the incentive to create new vacancies does not increase much. However, if wages would respond less, the increase in the share of income going to investors and the increase in employment would be bigger.

Recognizing the direct link between distribution and employment, several authors have proposed some mechanisms for generating smoother responses of wages and, therefore, larger fluctuations in income shares. Here we summarize three of these approaches. The first approach, proposed by Hagedorn and Manovskii (2008), is to assume that the flow utility received by workers when unemployed is not much smaller than the flow utility from working. In terms of the model, this is obtained by choosing a large value for the parameter $\overline{u}$, that is, the flow utility in the unemployment state. Although many consider this assumption implausible, the paper illustrates how this feature could bring the model closer to the data. The second approach proposed by Gertler and Trigari (2009) is to assume that wages are sticky. To illustrate these two cases, we first assign a higher value for the parameter $\overline{u}$ so that the replacement rate from unemployment is 95%. The impulse responses for this case, plotted in Figure 14.2, are labeled "High unemp. value." The figure also plots the impulse responses when the wage is exogenously fixed at the steady state with flexible wages (an extreme case of wage rigidity). As can be seen, both assumptions generate much higher volatility of employment and income distribution between investors and workers. This shows that inequality and macroeconomic volatility are closely interconnected: more volatile income distribution over the business cycle is associated with greater macroeconomic volatility. A third approach is explored in Duras (2013). The idea is that in periods with high productivity or output, the cost for workers to break the match is higher than in normal times. This weakens workers' bargaining position, alleviating the upward pressure on wages when productivity rises.

14.3.1.1.2 Shocks to Bargaining Power

Although it has been customary to assume that macroeconomic fluctuations are driven by productivity shocks, economic disturbances have many other possible sources. Here we summarize the effects of shocks that have a direct impact on the distribution of income, in particular, shocks that directly affect the bargaining power of workers, that is, the bargaining share η. When η decreases, a larger share of income will go to investors, increasing income inequality. At the same time, as investors appropriate a larger share of the surplus, they have a higher incentive to hire workers, thereby inducing a macroeconomic expansion. A similar approach has been studied by Ríos-Rull and Santaeulàlia-Llopis (2010) in the context of a neoclassical model.

14.3.1.1.3 Financial Shocks

The next step is to show that similar effects to those generated by shocks to bargaining power can be generated by the expansion and contraction of financial markets. The presentation of this case follows Monacelli et al. (2011).

Consider another slight modification of the search and matching model presented earlier, where we allow firms to borrow at the gross rate r. Borrowing, however, is subject to the constraint $b' \leq \varphi \mathbb{E} J'(b')$, where φ is stochastic. This variable captures the possible changes to the tightness of credit markets.

The firm enters the period with debt b. Given the new debt b' and the wage w, the dividends paid to investors are $d = z - w + b'/R - b$, where $R = (1+r)/(1-\lambda)$ is the gross interest rate paid by the firm conditional on survival. We assume that, in the event of exit, the firm defaults on the outstanding debt. Anticipating this, the lender charges the gross interest rate $R = (1+r)/(1-\lambda)$ so that the expected return from the loan is r. Notice that investors are shareholders and bondholders at the same time. We write the value functions exclusively as functions of debt, ignoring the potential variability of both productivity and bargaining power (that is, we now assume that z and η are constant).

The equity value of the firm can be written recursively as

$$J(b) = \max_{b'} \left\{ z - g(b) - b + \frac{b'}{R} + \beta(1-\lambda)\mathbb{E}J'(b') \right\} \tag{14.46}$$

$$\text{subject to} \quad b' \leq \varphi \mathbb{E} J'(b'),$$

where $w = g(b)$ denotes the (to be determined) wage paid to the worker. As we will see, the wage will depend on the debt. Notice that we have also used a prime to denote the next period value of equity, because this also depends on the next period aggregate states, specifically, the unemployment rate and credit market conditions. To avoid cumbersome notation, we do not include the aggregate states as explicit arguments of the functions defined here. Instead we use the prime to distinguish current versus future functions.

The value of an employed worker is

$$W(b) = g(b) + \beta \mathbb{E}[(1-\lambda)W'(b') + \lambda U'], \tag{14.47}$$

which is defined once we know the wage $g(b)$. The function U' is the value of being unemployed and is defined recursively as

$$U = \overline{u} + \beta \mathbb{E}[pW'(b') + (1-p)U'],$$

where p is the probability that an unemployed worker finds a job and $\overline{u}$ is the flow utility for an unemployed worker. Although the value of an employed worker depends on the aggregate states and the individual debt b, the value of being unemployed depends only on the aggregate states, because all firms choose the same level of debt in equilibrium. Thus, if an unemployed worker finds a job in the next period, the value of being employed is $W'(b')$.

The determination of the current wage solves the same problem as in Equation (14.38). We should take into account, however, that this solution also depends on b and on the function that determines future wages. Therefore, we write the solution of the bargaining problem as $w = \psi(g; b)$, where g is the function determining future wages. The equilibrium solution to the bargaining problem is the fixed point to the functional equation $g(b) = \psi(g; b)$.

Also in this case, the values for the firm and the worker satisfy

$$J(b) = (1-\eta)S(b), \tag{14.48}$$

$$W(b) - U = \eta S(b), \tag{14.49}$$

where the surplus is defined as $S(b) = J(b) + W(b) - U$. This can be written recursively as

$$S(b) = z - a - b + \frac{b'}{R} + (1-\lambda)\beta \mathbb{E}S'(b') - \eta\beta p \mathbb{E}S'(B'). \tag{14.50}$$

When a vacancy is filled, the newly created firm starts producing and pays wages in the next period. The only decision made in the current period is the debt b'. Therefore, the value of a vacancy just filled with a worker is

$$Q = \max_{b'} \left\{ \frac{b'}{1+r} + \beta(1-\eta)ES'(b') \right\} \tag{14.51}$$

$$\text{subject to} \quad b' \leq \varphi(1-\eta)\mathbb{E}S'(b').$$

Because the new firm becomes an incumbent starting in the next period, $S'(b')$ is the surplus of an incumbent firm defined in Equation (14.50). Notice that in the choice of b', a new firm faces a problem similar to that of incumbent firms (see problem 14.46). Even if the new firm has no initial debt and it does not pay wages, it will choose the same stock of debt b' as incumbent firms. We can then focus on a "representative" firm and in equilibrium $B = b$.

The value of posting a vacancy is equal to $V = qQ - \kappa$. Because in this version of the model there are no firm-specific productivity draws, $\kappa = \kappa_0 + \kappa_1$. As long as the value of a vacancy is positive, more vacancies will be posted. Free entry implies that $V = 0$ and in equilibrium we have

$$qQ = \kappa. \tag{14.52}$$

We now characterize the optimal choice of debt, that is, problem (14.46). Denoting by μ the Lagrange multiplier associated with the enforcement constraint, the first order condition is

$$\eta - R[1 + (1-\eta)\varphi]\mu = 0. \tag{14.53}$$

In deriving this expression, we have used the property of the model for which the choice of b' does not depend on the existing debt b and, therefore, $\frac{\partial S(b)}{\partial b} = -1$. We have also used the equilibrium condition $\beta R(1-\lambda) = \beta(1+r) = 1$.

Looking at the first order condition, we can see that the enforcement constraint is binding (that is, $\mu > 0$) if $\eta \in (0,1)$. Thus, provided that workers have some bargaining power, the firm always chooses the maximum debt and the borrowing limit binds. In this way, bargaining introduces a mechanism through which the financial structure is determined (Modigliani and Miller, 1958 does not apply). The reason is clear: by using outside finance, the firm is able to reduce the surplus that is bargained with the worker, increasing the possible rewards to equity.

To gather some intuition about the economic interpretation of the multiplier μ, it will be convenient to rearrange the first order condition as

$$\mu = \underbrace{\left(\frac{1}{1+(1-\eta)\varphi}\right)}_{\substack{\text{Total change}\\ \text{in debt}}} \times \underbrace{\left(\frac{1}{R} - \frac{1-\eta}{R}\right)}_{\substack{\text{Marginal gain}\\ \text{from borrowing}}} .$$

The multiplier results from the product of two terms. The first term is the change in next period liabilities b' allowed by a marginal relaxation of the enforcement constraint, that is, $b' = \varphi(1-\eta)\mathbb{E}S(b') + \bar{a}$, where $\bar{a} = 0$ is a constant. This is obtained by marginally changing $\bar{a}$. In fact, using the implicit function theorem, we obtain $\frac{\partial b'}{\partial \bar{a}} = \frac{1}{1+(1-\eta)\varphi}$, which is the first term.

The second term is the actualized net gain from increasing the next period liabilities b' by one unit (marginal change). If the firm increases b' by one unit, it receives $1/R$ units of consumption today in the form of additional dividends. In the next period, the firm has to repay one unit. However, the effective cost for the firm is lower than 1, because the higher debt allows the firm to reduce the next period wage by η, that is, the part of the surplus going to the worker. Thus, the effective repayment incurred by the firm is $1-\eta$. This cost is discounted by $R = (1+r)/(1-\lambda)$ because the debt is repaid only if the match is not separated, which happens with probability $1-\lambda$. Thus, the multiplier μ is equal to the total change in debt (first term) multiplied by the gain from a marginal increase in borrowing (second term).

Using the property for which the enforcement constraint is binding, that is, $\varphi EJ(b') = b'$, Appendix B shows that the wage can be written as

$$w = (1-\eta)\bar{u} + \eta(z - B) + \frac{\eta[p + (1-\lambda)\varphi]\kappa}{q(1+\varphi)}. \tag{14.54}$$

This equation makes clear that the initial debt B acts like a reduction in output in the determination of wages. Instead of getting a fraction η of output, the worker gets a fraction η of output "net" of debt. Thus, for a given bargaining power η, the larger the debt the lower the wage received by the worker. This motivates the firm to maximize the debt, as we have already seen from the first order condition.

Figure 14.3 plots the impulse responses to a credit shock, that is, a shock that raises φ and increases the credit available to firms. The credit expansion generates an increase in the capital income share and an increase in employment. Thus, changes in financial markets could alter the distribution of income and, with it, affect the incentives to create jobs. This is another example of how the distribution of income and macroeconomic performance are directly interconnected.

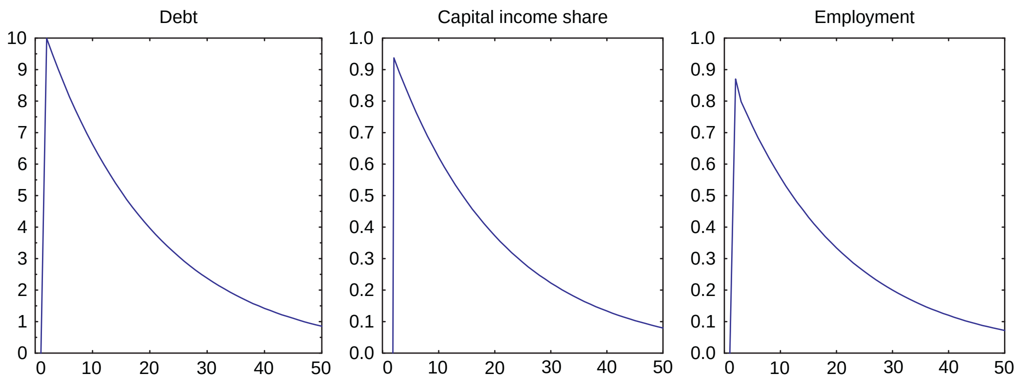


Figure 14.3 Impulse response to credit shock. The parameters are $\beta = 0.985$, $\alpha = 0.5$, $\mathbb{E}z = 1$, $\rho_z = 0.95$, $\sigma_z = 0.01$, $\bar{u} = 0.473$, $\kappa = 0.316$, $\lambda = 0.103$, and $A = 0.807$, $\overline{\varphi} = 0.0022$, $\rho_\varphi = 0.95$, $\sigma_\varphi = 100$.

14.3.1.2 Financial Accelerator and Inequality

A well-established tradition in macroeconomics introduces financial market frictions in business-cycle models. The key ingredients are based on two assumptions: market incompleteness and heterogeneity. Although not often emphasized, inequality plays a central role in these models. For example, in the seminal work of Bernanke and Gertler (1989) and (Kiyotaki and Moore (1997), entrepreneurial net worth is central to the amplification of aggregate shocks. When more resources are in the hands of constrained producers (i.e., these agents are richer), they can expand production and enhance macroeconomic activities. This can happen because they earn higher incomes or because their assets are worth more following asset price appreciations. Thus, these models posit a close connection between profit shares and the business cycle.

These models share some similarities with the matching models reviewed above: when a larger share of output goes to investors/entrepreneurs, the economy expands. At the same time, as the economy expands, a larger share of output or wealth (or both) is allocated to entrepreneurs. The mechanism of transmission, however, is different. In the matching model, the mechanism is the higher profitability of employment or investment. In the financial accelerator model, instead, it is the relaxation of the borrowing constraints. For a detailed review of the most common models used in the literature to explore the importance of financial frictions for macroeconomic fluctuations, see Quadrini (2011).[9]

[9] In addition to cyclical movements in the shares of income that go to labor and capital, there are cyclical movements in the shares of income earned by different groups of households. Castañeda et al. (1998) find that movements in unemployment rates by skill groups as well as movements in factor shares account for the bulk of the cyclical share of income earned by the various quintiles of households.

14.3.2 Low Frequency Movements in Inequality

We discuss here some theoretical ideas that have been proposed in the literature to explain some of the trends in the distribution of income that have occurred since the early 1980s. In Section 14.3.2.1 we look at the reduction in labor share, and in the following two sections we look at the increased inequality in wages and earnings. In Section 14.3.2.2 we examine the potential role of increased competition for human capital and in Section 14.3.2.3 the changes in the prices of skills due to skills-biased technical changes.

14.3.2.1 Labor Share's Reduction Since the Early 1980s

In a recent paper, Karabarbounis and Neiman (2014) document that labor share has significantly declined since the early 1980s for a majority of countries and industries. They pose a constant elasticity of substitution (CES) production function with nonunitary elasticity of substitution between labor and capital and argue that the well-documented decline in the relative price of investment goods (see Cummins and Violante, 2002; Gordon, 1990; Krusell et al., 2000) induced firms to substitute away from labor and toward capital. The consequence was the reduction in the price of labor. They conclude that roughly half of the observed decline in the labor share can be attributed to this mechanism.

To see how this mechanism works, consider the following aggregate production function:

$$Yt = F(K_t, N_t) = \left[\alpha_k (A_{K_t} K_t)^{\frac{\sigma-1}{\sigma}} + (1-\alpha_k)(A_{N_t} N_t)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}, \tag{14.55}$$

where σ denotes the elasticity of substitution between capital and labor in production, α_k is a distribution parameter, and A_{K_t} and A_{N_t} denote, respectively, capital-augmenting and labor-augmenting technology processes. As σ approaches 1, this becomes a Cobb–Douglas production function. Under perfect competition, marginal productivities yield factor prices and one can easily obtain expressions for labor share that crucially depend on the elasticity of substitution σ.

Karabarbounis and Neiman (2014) estimate σ by using only trends in the relative price of investment and the labor share in a cross section of countries. They find a value of 1.25. With this value, the decline in the (observable) relative price of investment accounts for 60% of the observed reduction in labor share. The remaining reduction can be imputed to larger increases in capital-augmenting technology relative to labor-augmenting technology, and to changes in noncompetitive factors (the most important, perhaps, are the permanent changes in the bargaining power of workers relative to firms along the lines of the models discussed earlier). Additional explanations arise from changes in the sectoral composition of output toward industries with higher capital share (which does not really seem to be the case), in part associated with globalization or the increase in the share of output that is traded with other countries, especially developing countries.

14.3.2.2 Increased Wage Inequality: The Role of Competition for Skills

There has been a big increase in earnings dispersion that is well documented in the literature. A glimpse of it can be seen in Tables 14.3 and 14.4. We now explore a possible explanation based on increased competition for skills in the context of human capital accumulation. We have already seen in Section 14.2.2.1 that human capital accumulation could be an important mechanism through which income becomes heterogeneous and that the higher the incentive to invest in human capital, the greater the degree of income inequality. We now look at one of the mechanisms that could affect the incentive to invest in human capital: competition for skills.

To study the importance of competition, we explore yet another version of the matching model described above by adding investment in human capital. We ignore here the possibility of firms to raise debt and limit the analysis to the version of the model without aggregate shocks. However, we now assume that output depends on the human capital of the worker, denoted by h. The production technology has a structure similar to the model presented in Section 14.2.2.1. The key feature of the model we look at now is that human capital investment requires both an output loss or pecuniary cost within the firm denoted by γ and a utility cost for the worker that we assume quadratic, $\frac{\alpha\gamma^2 h}{2}$. Given γ, human capital evolves stochastically according to

$$h' = h(1 + \gamma\varepsilon'),$$

where ε is an i.i.d. random variable. The gross growth rate of human capital is denoted by $g(\gamma, \varepsilon') = 1 + \gamma\varepsilon'$.

Because the outcome of the investment is stochastic, the model generates a complex distribution of human capital among workers. In the long run, the distribution will be degenerate because at the individual level, h follows a random walk. We assume that workers die with probability λ and that a match breaks down only when a worker dies. Thus, λ represents at the same time the death probability of a worker and the probability of separation of a match. In this way, the distribution becomes stationary and converges to a steady state.

There are contractual frictions that derive from the ability of the worker to control the investment γ after bargaining over the wage. The worker unilaterally chooses an investment γ that may be, and indeed is, different from the investment that maximizes the surplus of the match. This would be the investment that the worker would choose if he had been able to commit. Of course, when the firm bargains the wage, it anticipates the investment that the worker will choose in absence of commitment.

Let's define a few items. We can write the values of the investor and the worker in normalized form, that is, rescaled by human capital h. Then, the value for the investor can be written as

$$j = h - y - w + \beta(1-\lambda)\mathbb{E}g(y, \varepsilon')j', \tag{14.56}$$

where $j = J/h$ and w is the wage per unit of human capital. The total wage received by the worker is wh. The value for the worker is

$$\omega = w - \frac{\alpha y^2}{2} + \beta(1-\lambda)\mathbb{E}g(y, \varepsilon')\omega', \tag{14.57}$$

where $\omega = W/h$.

The value of being unemployed is

$$u = \overline{u} + \beta(1-\lambda)[p\omega' + (1-p)u'], \tag{14.58}$$

where $u = U/h$.

Even though in equilibrium, employed workers do not lose their occupation, u is important because it affects the threat value in bargaining. In a steady state we have $v = v'$, $\omega = \omega'$, $u = u'$.

The optimal investment y chosen by the worker maximizes the worker's value, that is,

$$\max_y \left\{ w - \frac{\alpha y^2}{2} + \beta(1-\lambda)\mathbb{E}g(y, \varepsilon')\omega' \right\},$$

with the first order condition given by

$$y = \beta(1-\lambda)\omega'\mathbb{E}\varepsilon'. \tag{14.59}$$

The important part to remember is that bargaining happens before the worker chooses her investment, which means that the surpluses that enter the problem take the investment y as determined by condition (14.59). From this condition we can see that y depends on ω' but not on the current value of ω, which implies that y is not affected by the outcome of the wage bargaining in the current period. Effectively, the current bargaining problem takes y as given and solves

$$\max_w \left\{ j^{1-\eta}(\omega - u)^{\eta} \right\}. \tag{14.60}$$

The first order condition implies that the parties split the net surplus, $s = j + \omega - u$, according to the bargaining weight η, that is,

$$j = (1-\eta)s,$$
$$\omega = \eta s + u.$$

As a comparison, we can also characterize the optimal investment when the worker commits to a particular y chosen to maximize the surplus of the match. In this case, the bargaining problem maximizes the objective (Equation 14.60) over both w and y. The first order condition with respect to w does not change, whereas the first order condition with respect to y becomes

$$1+\alpha\gamma=\beta(1-\lambda)(j'+\omega')\mathbb{E}\varepsilon'. \tag{14.61}$$

Compared with the optimality condition when the investment is controlled by the worker, Equation (14.59), we observe that the left-hand side and right-hand side terms in Equation (14.61) are both bigger. Therefore, the optimal choice of γ with commitment could be smaller or bigger. However, provided that α is sufficiently small, that is, the cost for the worker is not too large, the investment without commitment will be bigger.

14.3.2.2.1 General Equilibrium and the Impact of Competition

So far we have not worried about what happens outside the match, but there is a free entry condition that determines how many vacancies are posted. This is given by

$$q\beta j=\kappa, \tag{14.62}$$

where κ is the normalized cost of a vacancy.[10] One way of thinking about increased competition is that the entry cost κ is lower. We then have the following proposition.

Proposition 3.1 *The degree of competition κ affects the steady-state value of γ only in the environment without worker commitment.*

This result has a simple intuition: A lower κ is associated with a higher probability that an unemployed worker finds an occupation. As a result, the value of being unemployed increases. Inasmuch as this represents the threat value in bargaining, the worker can extract a higher wage w, which in turn increases the incentive to invest.

We can now show how an increase in competition (lower κ) affects inequality and aggregate outcomes simultaneously. In particular, we have that lower κ generates: (i) more risk taking and greater income inequality and (ii) higher aggregate income. The first effect can be seen from the first order condition (14.59). A lower entry cost increases the number of vacancies and, therefore, the value of finding another occupation if the worker quits. This allows the worker to bargain a higher wage, which in turn increases the employment value ω. We can then see from Equation (14.59) that a higher value of ω is associated with a higher γ. As we have seen in Section 14.2.2.1, a higher γ implies greater inequality. The second property—the increase in aggregate income—is obvious because a higher γ implies higher aggregate human capital. Thus, there is a trade-off between inequality and aggregate income.

Cooley et al. (2012) use a model with similar features but where the accumulation of human capital takes place in the financial sector. They show that greater competition for skills in the financial industry increased the incentive to invest in human capital and generated greater income inequality within and between sectors. This seems consistent with the recent increase in inequality, with income more concentrated at the very top of the

[10] For simplicity we are assuming that the cost of vacancies is proportional to the amount of human capital.

distribution and in certain professions, namely, managerial occupations in the financial sector. This pattern is also observed in the United Kingdom, as documented by Bell and Van Reenen (2010).

The idea that competition may increase inequality may go against the common wisdom that wealth is very concentrated because those who control wealth are able to protect it by limiting competition. From this the call is for increased enforcement of competition to reduce inequality. Of course, this does not mean that the theory described above is not valid. It depends on the particular environment we are studying: in certain sectors competition may lead to more inequality, in other sectors to lower inequality.

The degree of competition is just one way of affecting the equilibrium properties of aggregate income and inequality. Taxes are also important. In the context of this model, higher taxes discourage human capital investment (because the after-tax return from investing is lower), but this could be mitigated by the tax deductibility of the investment. Because the costs (curtailment of future earnings) and benefits (tax deductibility) occur at different stages of life when the individual has different incomes, the degree of progressivity becomes more important than the overall taxation. However, to the extent that taxes reduce investment, they also lower inequality (because lower investment reduces the volatility of individual incomes).

14.3.2.3 Skill-Biased Technical Change

In addition to increased competition for skills, which ends up rewarding those who are more skilled, a natural explanation for the increased earnings inequality is skill-biased technical change (Katz and Murphy, 1992). Although this term refers in general to changes in the distribution of earnings as a whole, it is often applied more specifically to the premium that college-educated people command compared with those without a college degree. This is motivated by the fact that the college wage premium, defined as the mean log wages of college graduates relative to high school graduates, has increased from 0.3 to 0.6 (see Goldin and Katz, 2009).

To illustrate how skill-biased technical change may have contributed to the increased earnings inequality, consider the following production function:

$$Y_t = F(A_{s,t}S_t, A_{u,t}U_t), \tag{14.63}$$

where S stands for the number of skilled workers (with a college education) and U stands for the number of uneducated workers (without a college education). $A_{s,t}$ and $A_{u,t}$ are exogenous technical coefficients that could change over time. Under perfect competition in the labor market, wages are marginal productivities, that is

$$w_{s,t} = A_{s,t}\frac{\partial F(A_{s,t}S_t, A_{u,t}U_t)}{\partial S_t}, \tag{14.64}$$

$$w_{u,t} = A_{u,t} \frac{\partial F(A_{s,t}S_t, A_{u,t}U_t)}{\partial U_t}. \tag{14.65}$$

The skill wage premium is defined as

$$\frac{w_{s,t}}{w_{u,t}} = \frac{A_{s,t}\dfrac{\partial F_t}{\partial S}}{A_{u,t}\dfrac{\partial F_t}{\partial U}}. \tag{14.66}$$

Absent large changes in the relative quantities of skilled and unskilled workers, we can assume that $\frac{\partial F_{t+1}}{\partial S} / \frac{\partial F_{t+1}}{\partial U}$ is very close to $\frac{\partial F_t}{\partial S} / \frac{\partial F_t}{\partial U}$, and the same goes for the marginal productivity of the unskilled. Consequently, we have that the change in the skill wage premium is given by

$$\frac{\dfrac{w_{s,t+1}}{w_{u,t+1}}}{\dfrac{w_{s,t}}{w_{u,t}}} \sim \frac{\dfrac{A_{s,t+1}}{A_{u,t+1}}}{\dfrac{A_{s,t}}{A_{u,t}}} = \frac{\dfrac{A_{s,t+1}}{A_{s,t}}}{\dfrac{A_{u,t+1}}{A_{u,t}}}. \tag{14.67}$$

This implies that the increase in the wage premium is owing to faster growth in the technology coefficient A_{s_t} relative to the growth of A_{u_t}, hence, the commonly used term "skill-biased technical change." But is there something more tangible than just an exogenous and largely unobserved technological change, or can we track it down to something observable?

Krusell et al. (2000) argued that we can relate these changes to something observable. Gordon (1990) and later Cummins and Violante (2002) have documented that the price of equipment (which is the main part of capital) in terms of consumption goods has gone down dramatically during the period of the rising skill premium. At the same time, the quantity of equipment has gone up significantly relative to output. This is a measurable form of technical change. Combine this with the notion that equipment or capital and skilled labor are complements, whereas unskilled labor is a substitute, and we have an actual channel through which technical progress is skill biased. The formulation in Krusell et al. (2000) does not have factor-specific technical change because all the effects of skill-biased technical change are in the increased quantity of equipment and can be written as

$$Y_t = K_t^{\alpha}\left[\mu U_t^{\sigma} + (1-\mu)\left(\lambda E_t^{\rho} + (1-\lambda)S_t^{\rho}\right)^{\frac{\sigma}{\rho}}\right]^{\frac{1-\alpha}{\sigma}}, \tag{14.68}$$

where K_t stands for structures (buildings) that play no role as they enter the production function in a Cobb–Douglas form, U_t and S_t are again unskilled and skilled labor, and E_t is equipment. Using observed measures of inputs, they estimated the elasticities of substitution ρ and σ and the share parameters α, λ, and μ, and found that unskilled labor is indeed a substitute for the aggregate of equipment and skilled labor, with both items being complementary to each other. They also found that this specification accounts very well for the observed wage premium under perfect competition for factor inputs.

Other forms of technical innovation indirectly generate skill-biased technical change. Suppose that technical change, regardless of its final effects on total productivity, is sometimes more dramatic than other change. The introduction of information technology could be one of these instances even if its impact on productivity is not as clear (Solow, 1987). Yet, the adaptation to this new technology may be easier for educated people. This is the approach taken by Greenwood and Yorukoglu (1974), Caselli (1999), and Galor and Moav (2000). Alternatively, suppose that information technology reduces information and monitoring costs within firms, allowing for reorganizations with fewer vertical layers and with workers performing a wider range of tasks. This gives educated workers an advantage. See, for example, Milgrom and Roberts (1990) and Garicano and Rossi-Hansberg (2004). Yet another form of skilled-biased technical change is an increase in competition for skills, as in the previous section, which could be the result of the technical change. In the context of the model studied earlier, the technical change can take the form of a lower vacancy cost κ. The lower κ increases the demand of skilled workers, which in turn increases the incentive to accumulate skills.

The technological innovations introduced in the 1970s seem to have affected the economy in other respects. Greenwood and Jovanovic (1999) and Hobijn and Jovanovic (2001) assume that new information technologies required a level of restructuring that incumbent firms could not face. As a result, their stock market value dropped. This is another form of redistribution in the sense that the owners of incumbent firms lost market value to the owners of new firms. Acemoglu (1998) has proposed a theory of the technical change itself being the result of a surge in college graduates.

The rise of superstars is another possible mechanism that increases the concentration of income. Rosen (1981) viewed the increase in earnings dispersion among people in some occupations as the result of an increase in their ability to reach more users of rare skills. Although this applies naturally to the case of artists and athletes, it also applies more generally to other types of skills. For example, Gabaix and Landier (2008) propose a theory of CEO pay where the value of managerial superstars is enhanced by the increase in the size of firms.

14.3.2.3.1 Skill-Biased Technical Change and Human Capital Accumulation

How does human capital investment interact with skill-biased technical change? Heckman et al. (1998) provide an answer that relies on the difference between observed wages and the price of skills that is due to the unpaid on the job investment of the Ben-Porath (1967) type models. They find no special role to capital in generating the increase in the skill wage premium. Instead, they find that the endogenous response of both more college attendance and the allocation of time to invest in further skills is sufficient to account for the patterns in the data. Guvenen and Kuruşçu (2010) also explore the interaction of skill-biased technical change and human capital accumulation emphasizing differences across people in the ability to acquire human capital. Guvenen and Kuruşçu (2010) argue that increased biased technical change immediately induces an increase in investment by talented individuals that first depresses the skill wage premium and then raises it and that is consistent with the observed bad performance of median wages and with the lack of increase in consumption inequality.

14.4. INEQUALITY AND FINANCIAL MARKETS

A large body of work links, theoretically, inequality and financial markets. The lack of complete markets helps to shape inequality through two channels. In Section 14.4.1 we study how the limited access to borrowing prevents poor households from undertaking valuable investments. This limited access keeps them and their descendants from climbing the social ladder. In Section 14.4.2 we study environments in which access to borrowing affects inequality, even when there are no household-specific investments.

In the environments studied in the first two sections, the borrowing limits are set exogenously. In Section 14.4.3 we start exploring endogenous theories of the borrowing limit by looking at environments in which the ability to borrow is limited by the incentive to default. In doing so, we follow the ideas suggested in Kehoe and Levine (1993). In Section 14.4.4 we review recent papers in which the limits to borrow come from the legal ability to default on debts allowed by the U.S. bankruptcy code. In Section 14.4.5 we explore various extensions of these models. Finally, in Section 14.4.6 we briefly discuss the literature that links the long-term performance of an aggregate economy with the ability of households to borrow.

14.4.1 Financial Markets and Investment Possibilities

Agents with available funds are not necessarily those with the best opportunities to use the funds. It is then socially desirable that the funds are channeled from the former to the latter, which is the primary role played by financial markets. Financial market imperfections, however, limit the volume of funds that can be transferred, and as a result the allocation is inefficient.

Financial market imperfections can take different forms. In a simple overlapping generations model in which the only decision that agents make is how much to invest in the education of their children, the lack of borrowing possibilities implies that investments with a rate of return higher than the risk-free rate will not be undertaken. A similar mechanism operates when there are borrowing possibilities but investments are risky and there are no insurance possibilities. This implies that investing agents may be left with very little consumption if they are unlucky. As a result, risk-averse agents may choose not to undertake investments.

To illustrate the importance of financial market frictions, consider an environment in which agents can save but cannot borrow, similar to that of Aiyagari (1994) developed in Section 14.2.1.3.2. The difference now is that the amount of efficiency units of labor is not random but is the result of investment. Consequently, two different investment strategies are available: households can save in the financial asset a (which in this model is backed by real capital), or they can invest in their own human capital so that $s' = \varphi(s, y)$, where y is the amount invested. The household's problem can be written as

$$v(s, a) = \max_{c, y, a'} u(c) + \beta v(s', a'; K), \tag{14.69}$$

$$\text{subject to} \tag{14.70}$$

$$c + a' + y = ws + a(1 + r), \tag{14.71}$$

$$s' = \varphi(s, y), \tag{14.72}$$

$$a' \geq 0. \tag{14.73}$$

If constraint (14.73) is not binding, the first order conditions of this problem imply

$$w\varphi_y(s, y) = 1 + r, \tag{14.74}$$

that is, the rate of return of the two types of investment is equalized. Moreover, imagine for simplicity that $\varphi(s, y) = \varphi(y)$; then in a steady state, all agents will have the same labor income. An interesting feature of this model is that convergence will arise immediately. That is, all households will have within a period the same labor income, because all agents will make the same investment in human capital. Differences in initial wealth perpetuate. For more general human capital production functions, we can get similar results, with the speed of convergence depending on the decreasing returns in s but not on y.

How would the analysis change when constraint (14.73) is binding? It all depends on the shape of function $\varphi(s, y)$. Let's start with the case $\varphi(s, y) = \varphi(y)$. Assuming that the function $\varphi(y)$ is strictly concave and $\varphi_y(y)$ approaches infinity as y approaches zero, all households will make some investment and the first order condition is

$$u_c(c) = \beta\varphi_y(y)u_c(c')w'. \tag{14.75}$$

This equation looks very similar to the Euler equation in the standard representative agent growth model, in which there is curvature in the production function. Consequently, no matter how poor they start out, all agents will slowly but steadily converge to a level of human capital that satisfies $\varphi_y(y)w = \beta^{-1}$. So the economy converges to equal labor income, even if the wealth distribution can be very unequal. Policies that subsidize investment in human capital could speed up the equalization process but will not change the eventual convergence outcome.

A lot of concern remains about *poverty traps*, that is, situations in which households that start with insufficient initial resources never abandon their poverty status. For this situation to happen, some special assumptions are needed. In particular, $\varphi(y)$ cannot be strictly concave. The typical assumption is to have a state of discontinuity such as a minimum expenditure that is needed to increase human capital. One example is the investment required for educational advancement. In this case, households compare the two options: whether to invest in education or not. If initial household wealth is very low, they may be unable to invest and still have positive consumption. But even households with slightly higher initial wealth may find that educational investment is feasible but not worth it, because it requires that initial consumption is way too low and the cost in utility terms too high. Clearly, in cases like these, government intervention can be fruitful because it is able to circumvent households' inability to borrow. The government can tax richer households today and transfer resources to the poorer households, or it can borrow, transfer resources to the poor households, and tax them later after they have acquired more education. A policy that makes education compulsory even at the cost of severe current disutility will not be optimal because poor households could have chosen to do it themselves if this choice were preferable.

Another possibility in which the structure of financial markets matters is to have a stochastic return to the investment technology. Consider a version of Equation (14.72) where higher investments in y yield a high expected value of s' but also a high variance. If the household had access to insurance markets, then it would happily undertake the investment, but if not, its risk aversion would prevent it from doing so. Again, in this case, certain government interventions that provide some form of insurance could be desirable.

14.4.2 Changes in the Borrowing Constraint

One way of assessing the role of financial constraints is to see what happens when they are relaxed. The Aiyagari (1994) model described in Section 14.2.1.3.1 assumes that financial markets are extremely underdeveloped: only one asset needs to be backed by physical capital, and there are no borrowing possibilities. What would happen if the financial constraints were to be relaxed, that is, what if we allowed for some noncontingent borrowing?

Table 14.7 shows the steady-state wealth distribution under various borrowing limits that go from a quarter of per-household yearly GDP to 1-year GDP. Figure 14.4 shows their associated Lorenz curves. We can see from the figure that, regardless of the calibration of the earnings process, inequality increases substantially with the relaxation of the borrowing constraint, in some cases to implausible levels (we cannot imagine an actual economy in which more than 60% of the population have negative financial assets). The Gini indices go up substantially, with one economy displaying a value above one, which is possible when we allow for negative values for the variable of interest. Looser borrowing constraints are associated with greater inequality because the poorest households want

Table 14.7 Distribution of wealth for various borrowing limits (in terms of per household yearly output)

Borrowing Constraint	Quintiles 1st	2nd	3rd	4th	5th	Top 10%	Top 5%	Top 1%	Gini
Low concentration of wealth economy (PSID)									
0.00	3.75	10.14	16.97	25.06	44.08	26.73	15.51	3.99	0.41
−0.25	2.39	8.82	15.21	24.10	49.48	30.90	17.81	4.44	0.46
−0.50	0.82	7.77	14.72	24.56	52.13	32.52	18.62	4.69	0.50
−1.00	−2.98	5.37	13.60	26.11	57.90	35.88	20.73	5.32	0.60
High concentration of wealth Aiyagari economy (SCF)									
0.00	0.00	0.00	0.50	4.78	94.72	76.96	51.77	14.54	0.86
−0.25	−1.72	−1.72	−0.23	3.56	100.11	81.47	54.83	15.50	0.93
−0.50	−3.50	−3.50	−0.88	2.89	104.99	85.16	57.03	16.19	0.99
−1.0	−7.21	−7.10	−1.87	2.88	113.30	90.75	60.68	16.94	1.11

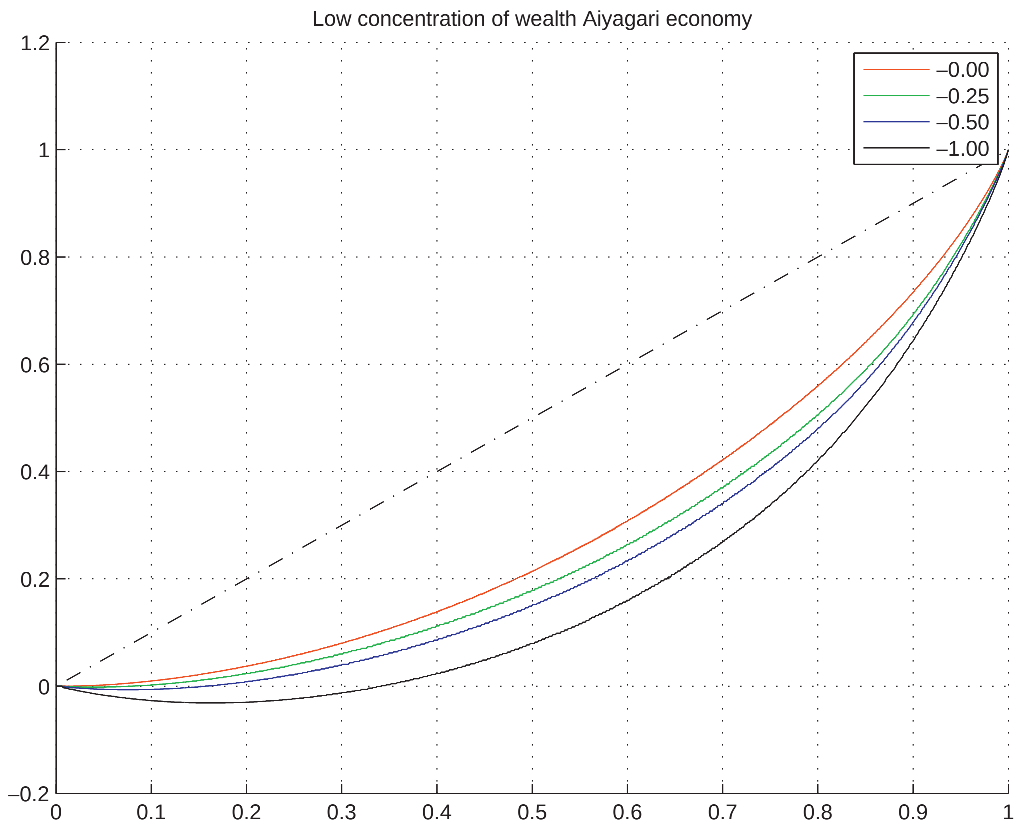
to borrow more. This result arises from the impatient nature of households in the general equilibrium. More specifically, households have a precautionary motive to save for the future when markets are incomplete, and on average they will never stop saving and will perpetually accumulate assets. But in a general equilibrium, the excess savings, which in aggregate takes the form of higher accumulation of capital, will drive down the marginal product of capital and, therefore, the return from savings. Consequently, in the steady-state equilibrium we have that $\beta^{-1} > 1 + r$, that is, households are more impatient than the return of their savings.[11] This result creates an incentive to anticipate consumption when the realization of earnings is low, which is made possible by the greater availability of credit (looser borrowing limit). This mechanism generates a greater concentration of wealth, as shown in Figure 14.4 and Table 14.7.

Another possibility is that improvements in the financial market allow agents not only to borrow more but also to buy insurance. In this case, agents could acquire assets or take liabilities with payments contingent on the realizations of idiosyncratic shocks. One

[11] To better understand why β^{-1} must be bigger than $1 + r$ in a steady-state general equilibrium, consider the following. Suppose that in a steady state $\beta^{-1} = 1 + r$. Given r, we can determine the stock of capital K from the equilibrium condition that equalizes the interest rate to the marginal product of capital. Lower interest rates must be associated with higher stocks of capital since the marginal product of capital is decreasing in K. Because agents face uninsurable risks, they save for precautionary reasons and, when $\beta^{-1} = 1 + r$, the average wealth accumulated by agents grows without bound (although individual wealth goes up and down stochastically, the average growth is positive). But in equilibrium the accumulated wealth is equal to K. Therefore, if wealth increases, K also increases, reducing the marginal product of capital and, with it, the interest rate r. As the interest rate declines, households save less until the average growth rate of wealth for the aggregate economy is zero. We therefore conclude that in a steady-state equilibrium, $1 + r$ must be lower than β^{-1}.

consequence is that households will no longer save for precautionary motives, because they can completely insure their individual consumption. Furthermore, there would not be too much aggregate savings, as in the economy without insurance. So the economy would slowly reduce savings until the interest rate became equal to the rate of time preference. Individual consumption could differ across households, as consumption depends on the initial distribution of physical and human wealth, as in Chatterjee (1994) (see Section 14.2.1.1).

Another important question is, how much borrowing can be sustained? In a model without leisure choice, the maximum sustainable debt is the one that can be paid in all states of nature. The worst state of the world is the lowest possible value of s, which we refer to as $\underline{s}$. A household that receives the lowest realization of earnings forever has the capability to pay a maximum amount of interest $\underline{s}w$. Thus, the maximum sustainable debt



(a)

Figure 14.4 Lorenz curves for various economies and borrowing limits. (a) Low concentration of wealth economy (PSID) calibrated as in Aiyagari (1994).

(Continued)

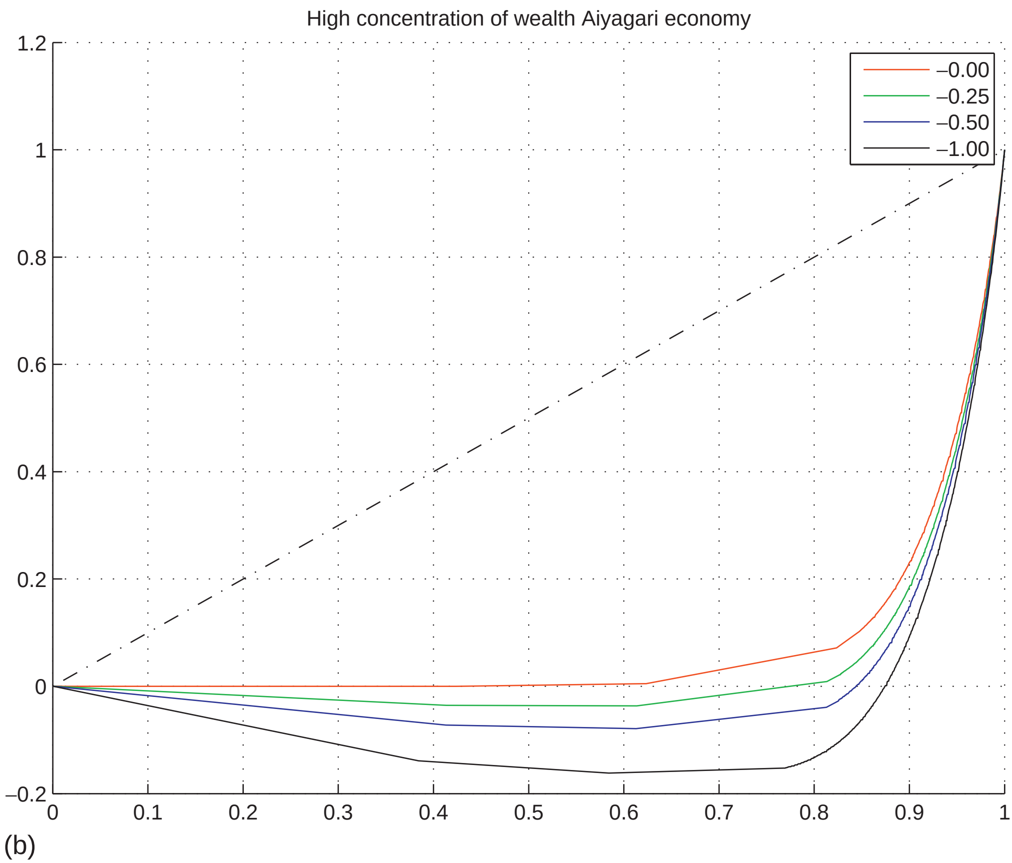


Figure 14.4—Cont'd (b) High concentration of wealth economy (SCF) calibrated as in Castañeda et al. (2003).

is $\frac{sw}{r}$, because the interest on this debt is exactly $\underline{s}$. Sometimes this is called the solvency constraint. Any debt larger than this value has a positive probability of not being paid.

14.4.3 Limits in the Ability to Borrow

So far, we have considered environments in which the access to credit markets is arbitrarily limited and there are no markets for contingencies. But why not? What limits the set of contracts that people can sign?

In a well-known and influential paper, Kehoe and Levine (1993) postulated that the ability of households to borrow is limited by their willingness to pay back when the alternative is to give up access to credit markets. In addition, a subset of the assets (physical or human or both) or endowments of the households can be seized, but not necessarily all of them. For example, future labor income may be outside of the reach of creditors.

Their approach does not preclude the existence of contingent markets. We would like to emphasize two features of this approach. The first feature is that it is the institutional environment that determines the set of contracts that are available. We think that this is an enormous advance compared with the literature that relies on exogenous borrowing limits. The second feature is that only the contracts that can be enforced ex post are available in the market. Therefore, once signed, there is complete compliance in the execution of these contracts. This feature is perhaps less appealing, as we see actual ex post reneging on formal contracts.

To show how this works, we could again slightly modify the Aiyagari economy. Let's first define the following object:

$$\overline{V}(s, a) = \max_{a' \geq 0} u(a(1+r) + sw - a') + \beta \sum_{s'} \Gamma_{s,s'} \overline{V}(s', a'), \tag{14.76}$$

which is the household's value without having access to borrowing. Moreover, with some abuse of notation, let's define $\overline{V}(s) = \overline{V}(s, 0)$ as the value attainable when the initial assets are zero. Clearly, this limit depends on the value of the shock s. Now consider the following problem:

$$v(s, a) = \max_{c, y, a'} u(c) + \beta_i v(s', a'; K), \tag{14.77}$$

$$\text{s.t.} \quad c + a' + y = ws + a(1+r), \tag{14.78}$$

$$a' \geq \underline{a}(s), \tag{14.79}$$

where $\underline{a}(s)$ is such that

$$v\left(s, \underline{a}(s)\right) = \overline{V}(s). \tag{14.80}$$

In words, households can borrow up to the level in which they would be better off in an autarkic state, that is, in a state in which they start from zero assets and never borrow again. Notice that it is quite possible that in this situation high-income people have more difficulties borrowing than low-income people, and this is because $\overline{V}(s^H) > \overline{V}\left(s^\ell\right)$ when $s^H > s^\ell$. Notice also that $\underline{a}(s)$ is an endogenous variable. We do not know its value before solving for the equilibrium of this economy.

We have written problem (14.77) under the implicit assumption that assets can never be confiscated. If the legal system were such that assets could be taken away in absence of compliance, we could substitute Equation (14.79) with

$$a' \geq \underline{\underline{a}}(s), \tag{14.81}$$

where $\underline{\underline{a}}(s)$ is such that $v\left(s, \underline{\underline{a}}(s)\right) = \overline{\overline{V}}(s)$ and $\overline{\overline{V}}(s) = u(sw) + \beta \sum_{s'} \Gamma_{s,s'} \overline{\overline{V}}(s')$. Essentially, the borrowing limit could be the amount that makes the agent indifferent between paying back the lender or defaulting and being forever unable to save or borrow.

In this model, all contracts are carried out, that is, loans and state-contingent contracts are always honored. In reality, however, many people file for bankruptcy. For example, in the 12 months between April 1, 2012, and March 31, 2013, 779,306 people filed for bankruptcy in the United States.[12] In some countries including the United States and Canada, debts are typically discharged after filing, whereas in other countries, such as Hungary, Romania, and Spain, there is no legal procedure to handle personal bankruptcy, and people are always liable for previous debts. The rest of the countries lie somewhere in between these extremes.

One possible strategy to deal with the pervasiveness of bankruptcy is to model it as a contingency fully negotiated ex ante by the parties. This strategy is hard to justify, however, because filing for bankruptcy is a legal procedure that can be completed unilaterally by the debtor. Hence, it is a right that cannot be forfeited. We need, then, to have explicit models that explicitly incorporate bankruptcy filings. One approach, followed within the optimal contracting tradition, is to assume that there are information asymmetries and costly state verification, as in Townsend (1979). The costly state verification model has been widely applied in macroeconomics, for example, in Bernanke and Gertler (1989), Bernanke et al. (1999), and Carlstrom and Fuerst (1995). In these models, default arises in equilibrium even if agents sign fully optimal contracts. In the next section, we will describe other approaches that are more in line with the literature that excludes the applicability of fully optimal contracts.

14.4.4 Endogenous Financial Markets Under Actual Bankruptcy Laws

During the last few years, considerable work has been done to bring together models with imperfect insurance and models with a legal system that allows agents to file for bankruptcy in a way that is similar to that of Chapter 7 in the U.S. bankruptcy code (Chatterjee et al., 2007; Livshits et al., 2007). We now present a version of these models and describe how their implications for the income and wealth distribution change compared with the basic Aiyagari model. These studies take advantage of a feature of the legal system that lists people who have filed for bankruptcy in public records for a certain number of years. The literature interprets the implications of the listing as limiting accessibility to borrowing for the duration of the public record.

Consider the following household problem, yet another variant of the basic Aiyagari problem:

$$v(s,a,1) = \max_{c,a'\geq 0} u(ws + a(1+r) - a') + \beta \sum_{s'} \Gamma_{s,s'} [(1-\delta)v(s',a',1) + \delta v(s',a',0)], \tag{14.82}$$

[12] See U.S. courts, http://www.uscourts.gov/.

$$v(s,a,0) = \max\left\{ u(sw) + \beta\sum_{s'}\Gamma_{s,s'}v(s',0,1), \quad \max_{c,a'} u(c) + \beta\sum_{s'}\Gamma_{s,s'}v(s',a',0), \right\} \tag{14.83}$$

$$\text{s.t.} \quad c + q(s,a')a' = ws + a. \tag{14.84}$$

The function $v(s, a, h)$ is the household's value function, with the last argument $h \in \{0,1\}$ denoting the household's credit history. When $h = 1$, the household's credit history is bad in the sense that the agent has defaulted in the near past and is prevented from having access to credit. Problem (14.82) depicts this case, showing that in the following period, its credit history may turn out to be good, $h = 0$, with parameter δ controlling the expected duration of market exclusion.

Problem (14.83) is of interest when the credit history is good, $h = 0$, and we have written it compactly, implicitly assuming that the household is in debt, $a < 0$. Here, the agent has two options: to file for bankruptcy or not. If the agent files, three things happen: household consumption equals current labor income, sw, credit history turns bad next period, $h' = 1$, and the household is prevented from saving. The latter property is a feature of the bankruptcy code, as the agent is not permitted to keep assets after filing for bankruptcy.[13] If the household does not file for bankruptcy, it can borrow or save as it wishes. Note, however, that we have written the budget constraint (14.84) differently from previous problems. The left-hand side, the uses of funds, has the asset position at the beginning of the following period multiplied by $q(s, a')$. This is the household-specific inverse of the interest rate. Lenders accurately forecast that the agent may file for bankruptcy and charge an extra premium so that in expected value they get the market return. The function $q(s, a')$ is an equilibrium object. If the household chooses to save, $a' \geq 0$, and the inverse of the interest rate is that of the safe asset: $q(s, a') = (1 + r)^{-1}$.

The optimal solution to the problem of a household with negative assets is to default for a range of its earnings. The set of earnings for which the household defaults increases with the stock of debt.

The solution to this problem has two interesting properties. First, inasmuch as default is costly (the household will not be able to borrow for a while), the household would not default if the debt is very small. Second, in some circumstances, the household may be too poor to default, opting instead to borrow even more for a sufficiently low realization of earnings. Consequently, the equilibrium of this model requires that the inverse of the interest rate $q(s, a')$ is such that lenders break even in expected value, which in turn implies that interest rates are increasing in the amount borrowed and in the likelihood of bad earnings realizations.

[13] In the United States, the agent can keep a maximum amount of assets, and this amount varies across states. Here, we have assumed zero retainable. In the discussion from which we are abstracting are other subtleties of the bankruptcy code, such as the requirement that labor income is below the state median's income.

This structure has proved useful in making sense simultaneously of the extent of unsecured borrowing in the United States as well as the frequency of bankruptcy filings, especially if the model is enhanced with a few bells and whistles such as expenditure shocks.

14.4.4.1 A Weakness of This Approach

Why are households refused credit when they have a bad credit history? Nothing in the law requires this. In fact, the opposite is true: in the United States, a bankruptcy filing under Chapter 7, the relevant case, precludes additional filings over a period of years, making a recent filer appear to be a better creditor than somebody with a clean slate.

This question has three possible answers, but none are completely satisfactory. First, a Nash equilibrium with a coordination problem can be constructed when lenders believe that agents with bad credit histories will not pay back and hence they will not lend, whereas prospective borrowers might as well choose to default, because they do not receive credit. Although this is indeed a Nash equilibrium, it is one that is always present in the event of lending, and there is no argument for why it happens only with a bad credit history. Another possibility is to construct a trigger strategy whereby lenders coordinate not to lend during a punishment period in the event of default. But like all triggers, this is not an equilibrium of the limit of finite economies. Hence, it is not a Markov equilibrium. Many economists are comfortable with trigger strategy equilibria, whereas others are not. The last rationale for exclusion of those with bad credit is to postulate the existence of a regulator that prevents lenders from lending to those with bad credit, something that is not actually done by any of the banking regulators.

14.4.5 Credit Scoring

Chatterjee et al. (2008) and Chatterjee et al. (2004) propose a solution to the weakness of models based on exogenous exclusion after bankruptcy filings. These papers note that in the United States, there is pervasive use of credit scores, which are assessments of reliability made by independent companies. The authors then pose a model in which two types of people differ in some fundamental attribute associated with reliability that is not directly observable by outsiders—for instance, patience or even good driving habits. The credit score is then used as the market assessment of being a good type, meaning the type that is more likely to pay back debts or be reliable. In this context, both types of agents fall under the model of borrowing in which there are multiple types. The key here is that both types of agents—both the patient and the impatient agents—want to repay their debts to signal that they are patient, which allows them to have access to better borrowing terms. In this context, filing for bankruptcy increases the market-assessed likelihood that an agent is of the bad type, which translates to a severe worsening of loan terms, if not an outright exclusion of future credit. Moreover, because the market is assessing traits that are relevant not only for the repayment of credit but also for other things (e.g., cheap property insurance, access to rental property, personal relationships), timely repayment of debts carries a strong incentive that allows for the possibility of many contracts to be carried out, even if the law lacks the necessary teeth to enforce these contracts.

14.4.6 Financial Development and Long-Run Dynamics

We now look at the extent to which access to financial markets can help us to understand the long-run dynamics of the economy. In Section 14.4.6.1 we briefly focus on long-term growth, and in Section 14.4.6.2 we discuss how the evolution of financial markets can also help us understand the issue of global imbalances, that is, the emergence of large and persistent balance of payments deficits.

14.4.6.1 Long-Run Growth and Financial Development

The Schumpeterian view places entrepreneurship at the center stage of economic development. Owing to financial constraints and the lack of insurance markets, however, entrepreneurial investment is suboptimal. Essentially, when financial markets are not well developed, resources cannot be redistributed from those who control the resources but do not have the best uses of these resources to those who have the best investment opportunities but lack the funds. This efficiency problem is especially severe when the distribution of resources is particularly concentrated. We may then end up with a situation in which the poor become (relatively) poorer because they cannot take advantage of investment opportunities and the economy as a whole grows less. Examples of studies that emphasize the importance of inequality for growth in the presence of financial constraints are Galor and Zeira (1993), Banerjee and Newman (1993), and Aghion and Bolton (1997). Because these studies were already reviewed by Bertola (2000) in a previous edition of the Handbook, we do not repeat their description in this chapter.

A more recent literature, however, also emphasizes that market incompleteness—that is, environments in which the trade of state-contingent claims is limited—could have both positive and negative effects on capital accumulation. In a world with only uninsurable and exogenous earning shocks as in Aiyagari (1994), market incompleteness generates more capital accumulation and, therefore, more growth. When risky income is endogenous, however, as in Angeletos (2007), market incompleteness may discourage investment. See also Meh and Quadrini (2006).

Another group of studies that investigates the relation between inequality and macroeconomic performance emphasizes the importance of social conflict and expropriation. Greater inequality often associated with underdeveloped financial markets means that a larger group of individuals are at the bottom of the distribution and face poor economic conditions compared with the rest of the population. Faced with poor economic conditions and the feeling that the prospects for economic improvement are impaired by the excessive concentration of wealth, the resentment toward the rich starts to rise, which creates incentives to expropriate either by stealing or through revolutions. The risk of expropriation has two negative effects. First, it acts as an investment tax that discourages investment. Second, agents devote more resources to protect property rights instead of using the resources for productive and growth-enhancing activities. Benhabib and Rustichini (1996) develop a model that formalizes this idea. Although not explicitly considered in this paper, financial underdevelopment could contribute to this because it makes it more difficult for poor people to escape from poverty.

Another theory of inequality affecting growth is developed in Murphy et al. (1989). This paper assumes that some technologies have increasing returns. These technologies become profitable only if the domestic market is sufficiently large, that is, enough demand exists for the goods produced with the new technologies. If wealth is highly concentrated, the domestic market remains small (because not enough consumers can afford these goods). As a result, growth-enhancing technologies will not be implemented. The paper does not explicitly explore the role of financial markets; however, to the extent that financial underdevelopment creates the conditions for greater concentration of wealth, the mechanism described in this paper becomes more relevant in economies in which the financial structure is relatively underdeveloped.

Kumhof and Rancière (2010) have proposed an explanation for the recent crisis based on the changes in income distribution pinpointing similarity with the Great Depression. The idea is that, because of an exogenous shock that affected the ability of the rich to grab earnings, income became more concentrated, and as a result, the poor started to borrow more, increasing the debt-to-income ratio in the economy. Eventually, the increase in borrowing triggered the crisis.

We conclude this section by citing the work of Greenwood and Jovanovic (1990). Although this paper does not deal directly with inequality, it shows that improvements in financial markets (in this particular case, through the information gathered by financial intermediaries) have important effects on economic growth. As we have seen in previous chapters, market incompleteness also creates inequality. Therefore, once complemented with the previous analysis, this paper could also be relevant for understanding the link between financial market development, inequality, and growth. Also important is the work of Greenwood et al. (2010).

14.4.6.2 Global Imbalances

We have not talked much about cross-country inequality because this topic is usually a concern for development-oriented economists. However, inequality may be shaped by the increase in trade that is properly known as globalization, which is due to the reduction in the trade barriers for both technological and policy reasons. We have already referred, if only obliquely, to a mechanism by which more trade across countries could affect inequality: opening to trade changes the relative price of skills and may be behind part of the recent increase in the wage-skill gap. But an increase in trade shapes inequality both within and between countries through other, more subtle mechanisms. In this section we illustrate some potential mechanisms through which inequality is linked to globalization. Further analysis of the role of globalization for inequality is conducted in Chapter 20.

The process of international globalization is commonly presented as taking the form of higher trade in goods and services (imports and exports) as a fraction of GDP. But there is another side to it. Several advanced countries, the United States in particular, have experienced over the last 30 years a persistent deficit in the balance of payments as a result of imports being higher than exports, with the consequent deterioration in their net foreign asset positions. On the other hand, oil-producing countries and several emerging countries, China in particular, have been accumulating positive net foreign asset positions. "Global imbalances" is the term often used to refer to the situation in which some countries accumulate large negative net foreign asset positions whereas others accumulate positive net foreign asset positions. This situation has affected inequality, but to understand the impact on inequality we first need a theory of why imbalances could emerge in the wave of globalization.

Mendoza et al. (2007) provide one such theory. They claim that sustained deficits cannot be explained solely with traditional trade forces (different factor prices, technological advantages, or lower transportation costs). We also need to understand the differential saving behavior of countries that in equilibrium lead to different rates of returns on savings (insofar as international financial markets are somewhat segmented). This is possible even if countries have identical preferences and production technologies, but agents in each country differ in the extent to which they are capable of insuring their individual risks. This can be illustrated with the now familiar Aiyagari economy.

Suppose that we compare two economies, both slightly modified versions of the Aiyagari environment described above, that differ only in the process for earnings, one being more volatile than the other. It is important to point out that the assumption that countries differ in the volatility of earnings is a shortcut to capture other, more micro-founded differences. For example, in Mendoza et al. (2007), countries do not differ in the underlying process for earnings but in the sophistication of financial markets. Agents (consumers and firms) in countries with more advanced financial markets have a better opportunity to insure their idiosyncratic risk. Because in terms of savings the implication of higher insurance is similar to lower variability of earnings, here we illustrate the mechanism by assuming lower earning volatility. In some applications the higher ability to insure could derive from government policies (for example, the provision of public-funded health insurance). In some cases, the differences could come from more uncertainty about the underlying process for earnings. For example, a country that is experiencing a process of transformation (such as China, during the last three decades) is also possibly characterized by greater uncertainty at the individual level. Independently of the actual sources (greater ability to insure or greater underlying uncertainty), it should be clear that the example provided here is just a shortcut to illustrate something more fundamental such as differences in the characteristics of the financial system.[14]

[14] See Mendoza et al. (2007) for more details on how differences in the financial system can lead to lower ability to insure.

Table 14.8 Two economies before and after being able to borrow from each other

	Before: autarky		After: mobility	
Economy	**Low var**	**High var**	**Low var**	**High var**
Capital to output ratio	3.34	3.88	3.67	3.67
Interest rate (%)	4.02	1.27	2.24	2.24
Wealth to output ratio	3.34	3.88	0.39	6.95
Gini index of wealth	0.41	0.59	0.50	0.39
Coeff. of var wealth	0.76	1.09	0.88	0.96
1st quantile	3.40	0.00	0.40	0.00
2nd quantile	10.21	0.69	6.06	3.35
3rd quantile	17.11	15.50	17.58	17.45
4th quantile	25.16	30.21	25.94	21.00
5th quantile	44.15	53.60	50.01	58.20
Top 10% (cumulative)	26.73	33.28	30.37	29.63
Top 5% (cumulative)	15.51	19.72	18.09	17.21
Top 1%	3.99	5.19	4.25	4.32

To this end, we use two different processes for earnings. The first process is what we used above for the version of the model that we called PSID economy or low-variability economy. The second process is what we used in the high-variability economy, a version of the SCF economy with a slightly less extreme good state. Besides the process for earnings, the two economies are alike in all other dimensions.

The first two columns of Table 14.8 display the steady states of these two economies under autarky. The first column, the low-variability economy, has a capital-to-output ratio of 3.34, implying an annual interest rate of 4.02%. Because in this economy wealth can only take the form of capital, total wealth is also 3.34 × output, and this is what households choose to hold to accommodate the shocks to earnings given the 4.02% interest rate. The second column of Table 14.8 refers to the economy with higher income variability also in the autarky regime. Households choose to hold more wealth (3.88 × output) to bear the high risk. Two things to note are that the interest rate is now much lower, 1.27%, and that output is slightly higher because of the higher capital.

The determination of the equilibrium is depicted in panel (a) of Figure 14.5. This figure plots the aggregate (steady-state) supply of savings as an increasing, concave function of the interest rate.[15] The demand for savings is downward sloping because of the diminishing marginal productivity of capital. Country 1 has a lower volatility of

[15] Aggregate savings converge to infinity as the interest rate approaches the rate of time preference from below, because agents need an infinite amount of precautionary savings to attain a nonstochastic consumption profile.

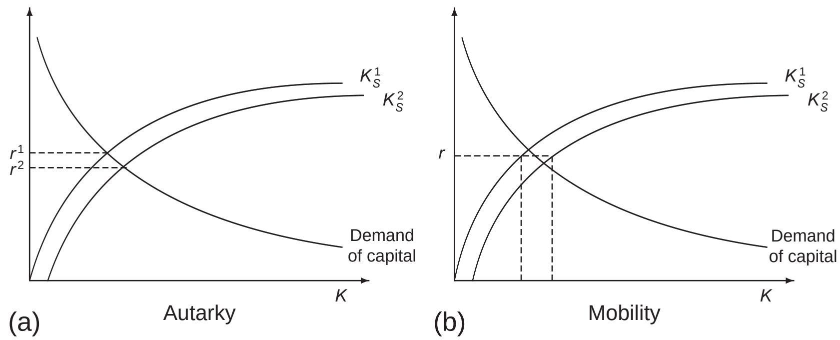


Figure 14.5 Steady-state equilibria with heterogeneous earning risks.

individual earnings and hence a lower supply of savings for each interest rate. As a result, the equilibrium in autarky implies a higher interest rate and a lower total capital.

Imagine now that households in these two economies can start owning capital in the other country, that is, the countries become financially integrated. After a brief period of transition that depends on the ease with which physical capital can flow or be reallocated, the interest rate in both countries will be equalized. This implies that the low-variability economy will experience a reduction in the interest rate and the high-variability economy will experience an increase in the interest rate. Then, in the country in which the interest rate decreases (low-variability economy), savings will fall, whereas in the country in which the interest rate increases (high-variability economy), savings will rise. The result is that households in the high-variability economy end up owning part of the capital installed in the low-variability economy. In this way, global imbalances may emerge as the low-variability economy dis-saves. Effectively, the low-variability economy consumes and invests more than it produces, with the difference covered by imports in excess of exports (trade deficit).

This process takes a long time until the aggregate savings of households in each country no longer change. This new steady state is depicted in panel (b) of Figure 14.5. The world interest rate is somewhere between the pre-liberalization interest rates in the two countries. Compared with autarky, the interest rate and the supply of savings fall in country 1 and rise in country 2, and hence the country with lower volatility of earnings ends up with a negative foreign asset position. Moreover, the capital stock rises relative to its autarky level in country 1 and falls in country 2. Thus, financial globalization leads capital to flow from economies with more risk to those with lower risk.

For analytical simplicity, we have modeled this process as the outcome of countries that differ in their earnings risk. However, as emphasized above, this is just a shortcut to capture other types of differences across countries that ultimately lead to different exposure to risk. It could very well be the case that the underlying risk is identical across countries but that the lower risk in one country is just the result of more-developed financial markets, which allow for higher insurability of risk. Formally, in the first country there are more markets for state-contingent claims. This is the approach taken in Mendoza et al. (2007), and the end result is similar to the case of differential processes for earnings: the country with a higher ability to insure saves less and has a higher interest rate than the country with less-developed financial markets. When the two countries integrate, it is the more financially developed country that accumulates negative foreign assets, whereas the less financially developed country accumulates positive foreign assets.

Therefore, financial market differences can affect the distribution of wealth across countries: in the long run, countries that are more financially sophisticated become poorer relative to countries that are less financially sophisticated (compared to the pre-liberalization era). This, however, does not mean that liberalization is welfare reducing for developed countries and welfare improving for less-developed countries. In Mendoza et al. (2007) we found, somewhat surprisingly, that liberalization was welfare improving for developed countries but slightly welfare reducing for less-developed countries (based on an equally weighted welfare function). In our example displayed in Table 14.8, the international redistribution of wealth is quite large, with the low-variability country ending up with barely 5% of total wealth. Yet, it started with almost half. The large international redistribution of wealth follows from the assumption that there are large differences in risk between the two countries. In reality, especially among integrated countries, the differences in risk may not be that big. Also, when a country accumulates too many foreign liabilities, there could be an incentive to default on these liabilities. This imposes a limit on the redistribution of wealth that can be generated across countries through this mechanism. Nevertheless, this example suggests that differences in savings could generate significant inequality in wealth across countries.

Cross-country financial market heterogeneity also plays a central role in Caballero et al. (2008) for explaining global imbalances. The mechanism proposed in this paper does not rely on risk but on the availability of saving instruments. The idea is that in certain countries, savers have difficulty storing their savings in high return assets. The implications for global imbalances, however, are similar to Mendoza et al. (2007). The two mechanisms are complementary ways of thinking about how the characteristics of financial systems can shape the distribution of wealth across countries in a globalized world. Interestingly, these contributions illustrate another mechanism through which financial globalization redistributes wealth. When productive inputs are not perfectly reproducible (as in the case of land), liberalization also leads to the equalization in the prices of these assets. Because under autarky these assets were cheaper in financially developed countries, these countries experience capital gains, whereas countries with less-developed financial markets experience capital losses.

The process of international redistribution of wealth also has consequences for the internal wealth distribution within each country. We see how wealth concentration increases in the low-variability country as measured by either the Gini index of the coefficient of variation of wealth or even by the shares held by the richest households (see Table 14.8). The opposite process happens in the less financially developed country, where the wealth distribution becomes more equal after international financial integration. Perhaps this process has contributed, at least in part, to the increased wealth concentration in the United States that we documented earlier.

14.5. THE POLITICAL ECONOMY CHANNEL

We have already seen in the previous sections some channels through which the distribution of income and wealth is interconnected with the aggregate performance of the economy. In this section, we discuss one particular channel through which inequality affects economic activities, that is, through the political and institutional system. Because many policies have redistributive consequences, the degree of inequality plays a central role in the choice of policies because societies with more unequal distributions of resources might demand greater redistribution. Because redistributive policies are often distortionary, the result is that more unequal societies tend to experience lower income or growth (or both).

Many contributions emphasize this mechanism, starting with Meltzer and Richard (1981). Examples are Persson and Tabellini (1994), Alesina and Rodrik (1994), Krusell and Ríos-Rull (1996), and Krusell et al. (1997). Many of these contributions, however, ignore individual uncertainty, which in a dynamic environment could play an important role in affecting the demand for redistribution as well as the distortions associated with redistributive policies. The goal of this section is to present a simple framework that illustrates the central idea of the early literature. It shows how the consideration of idiosyncratic uncertainty enriches the analysis and makes the relation between inequality and redistribution more complex than in these early studies.

14.5.1 A Simple Two-Period Model

Suppose that there is a continuum of agents who are alive for two periods. Agents value consumption, c_t, but dislike working, h_t, according to the utility function

$$u\left(c_t - \frac{h_t^2}{2}\right).$$

There are two sources of income: endowment, η_t, and labor, h_t. Individual endowments evolve according to

$$\ln(\eta_{t+1}) = \rho \cdot \ln(\eta_t) + \varepsilon_{t+1},$$

where $\varepsilon_{t+1} \sim N(0, (1-\rho^2)\cdot\sigma^2)$. This implies that the economywide distribution of log-endowments is normal with mean zero and variance σ^2, that is, $\ln(\eta_t) \sim N(0, \sigma^2)$. By changing ρ we change the persistence of endowments, but we keep the economywide distribution (inequality) constant. This parameter determines the degree of mobility: higher values of ρ imply lower mobility.

Before continuing, it will be helpful to derive some of the key moments of the cross-sectional distribution of endowments. Because endowments are log-normally distributed, that is, $\eta \sim LN(0, \sigma^2)$, the mean and the median are, respectively,

$$\mathrm{Mean}(\eta_t) = e^{\frac{\sigma^2}{2}}, \quad \mathrm{Median}(\eta_t) = 1.$$

These are unconditional moments. Also convenient is to derive the expected next-period endowment for an individual with current endowment η_t^i. The conditional mean is

$$\mathbb{E}\left[\eta_{t+1}^i | \eta_t^i\right] = e^{\rho \ln\left(\eta_t^i\right) + \frac{(1-\rho^2)\sigma^2}{2}}.$$

This conditional expectation will play an important role in the analysis of the model. Here we observe that if $\rho = 0$, the expected next-period endowment is the same for all agents. For an agent with median endowment, that is, $e_t^m = 1$, the conditional expectation becomes

$$\mathbb{E}\left[\eta_{t+1} | \eta_t^m\right] = e^{\frac{(1-\rho^2)\sigma^2}{2}}.$$

We can then compute the ratio of the next period economywide average endowment over the next period endowment expected by an agent whose current endowment is the median value. This is equal to

$$\frac{\mathrm{Mean}(\eta_{t+1})}{\mathbb{E}\left[\eta_{t+1} | \eta_t^m\right]} = e^{\frac{\rho^2\sigma^2}{2}}.$$

This expression makes it clear that the difference between the average endowment and the endowment expected by the agent with the median endowment in the current period depends on the persistent parameter ρ. The difference becomes zero if there is no persistence, that is, $\rho = 0$, and it is maximal when $\rho = 1$. Although the parameter ρ affects the ratio between the average endowment and the expected endowment by the median agent, ex post inequality does not depend on ρ. In fact, we have that

$$\frac{\mathrm{Mean}(\eta_{t+1})}{\eta_{t+1}^m} = e^{\frac{\sigma^2}{2}}.$$

We will use these moments below, after completing the description of the model.

The government taxes incomes, from endowment and labor, at rate τ_t and redistributes the revenues as lump-sum transfers. The budget constraint for the government is

$$T_t = \tau_t \int_i \left(\eta_t^i + h_t^i\right) di,$$

where i is the index for an individual agent.

Agents do not save and solve a static optimization problem. Given the tax rate and the transfer, an individual agent i maximizes the period utility by choosing the labor supply h_t^i, subject to the following budget constraint:

$$c_t^i = \left(\eta_t^i + h_t^i\right)(1-\tau_t) + T_t.$$

Taking first order conditions with respect to h_t for an individual worker with endowment η_t^i, we get the supply of labor $h_t = 1-\tau_t$. Substituting in the utility function and using the equation that defines the government transfers, we get the indirect utility for period t:

$$U^i(\tau_t) = u\left(\tau_t \int_\eta \eta dF_t + \tau_t(1-\tau_t) + \eta_t^i(1-\tau_t) + \frac{(1-\tau_t)^2}{2}\right).$$

Now suppose that agents vote for the next period tax rate τ_{t+1}. The tax rate preferred by an agent with current endowment η_t^i maximizes the expected next period indirect utility, that is,

$$\max_{\tau_{t+1}} \mathbb{E}_t\left[u\left(\tau_{t+1}\int_\eta \eta dF_{t+1} + \tau_{t+1}(1-\tau_{t+1}) + \eta_{t+1}^i(1-\tau_{t+1}) + \frac{(1-\tau_{t+1})^2}{2}\right)\eta_t^i\right],$$

where we have denoted by $F(\eta)$ the distribution of endowments. Because the log-endowments are normally distributed, $F(\eta)$ is a log-normal distribution.

Notice that the voter forms expectations about the future endowment conditional on the current endowment. Of course, the higher is the persistence, the higher the dependence of the expected value from the current value.

Taking the first order condition, we derive

$$\tau_{t+1}^i = \int_\eta \eta dF_{t+1} - \mathbb{E}\left[\eta_{t+1}^i|\eta_t^i\right] - \frac{\mathrm{Cov}\left(dU_{t+1}^i, \eta_{t+1}^i|\eta_t^i\right)}{\mathbb{E}\left[dU_{t+1}^i|\eta_t^i\right]}, \tag{14.85}$$

where dU_{t+1}^i denotes the derivative of the indirect utility for agent i with respect to the next-period tax rate. Notice that this term also depends on the tax rate. The above condition implicitly determines the tax rate.

The first term on the right-hand side of Equation (14.85) is the mean value of the economywide endowment, which is equal to $e^{\frac{\sigma^2}{2}}$. This term is the same for all agents. The second term is the expected endowment of agent i given the current endowment. This term is increasing in η_t^i, unless $\rho \leq 0$, which is excluded by assumption. Therefore, ignoring the third term, the preferred tax rate decreases with the current endowment.

The third term captures the role of risk aversion. Because the utility function is strictly concave and its derivative is strictly decreasing, $dU_{t+1}^i(.)$ decreases with the realization of next period endowment η_{t+1}^i, implying that the covariance term is negative. Therefore, preferences for taxes increase with the concavity of the utility function. This is the effect of risk aversion.

14.5.1.1 The Case of Risk Neutrality

Because the third term in the first order condition (14.85) is itself a function of τ^i_{t+1}, it is difficult to derive an analytical expression for the tax rate. Therefore, we first specialize to the case with risk-neutral agents so that $\text{Cov}(dU^i_{t+1}, \eta^i_{t+1} | \eta^i_t) = 0$ and the preferred tax rate reduces to the first two terms in Equation (14.85). We can then establish that the preferred tax rate is monotonically decreasing in the current endowment η^i_t and the equilibrium tax rate is the one preferred by the agent with the median endowment. Using the fact that endowments are log-normally distributed and the log-endowment of the median voter is zero, the conditional expectation of the median voter for the next period endowment is $\mathbb{E}\left[\eta_{t+1} | \eta^m_t\right] = e^{\frac{(1-\rho^2)\sigma^2}{2}}$, whereas the economywide average is $\int_\eta \eta dF_{t+1} = e^{\frac{\sigma^2}{2}}$. Substituting in the preferred tax rate, we obtain the equilibrium tax rate

$$\tau_{t+1}\left(e^m_t\right) = e^{\frac{\sigma^2}{2}} - e^{\frac{(1-\rho^2)\sigma^2}{2}}. \tag{14.86}$$

The first two terms capture the standard politico-economy theory: because the average endowment, $e^{\frac{\sigma^2}{2}}$, is bigger than the median endowment, $e^{\frac{(1-\rho^2)\sigma^2}{2}}$, there is demand for redistribution. If we increase inequality by raising σ, the demand for redistribution increases. Because the optimal effort chosen by all agents is $h = 1 - \tau$, higher taxes discourage effort with negative effects on aggregate production. In some of the models proposed in the literature, taxes distort the accumulation of capital instead of effort, but the idea is similar.

The mechanism described above links inequality to redistribution and macroeconomic activity and captures the key features of the model studied in Meltzer and Richard (1981). In addition to this mechanism, the model presented here emphasizes the role of mobility captured by the parameter ρ. If we reduce ρ so that the economy experiences higher mobility, the cross-sectional inequality does not change. In fact, the ratio of average endowment and median endowment remains $e^{\frac{\sigma^2}{2}}$. However, the tax rate preferred by the median voter declines, as we can see from Equation (14.86). Even if the median voter has low endowment in the current period, what matters for next period taxes is the future endowment. If mobility is high, the median voter does not expect to keep the low endowment in the future. Thus, it is not optimal to choose high tax rates. In the limiting case with $\rho = 0$, the expected future endowment for all agents will be the average endowment and, in expected terms, the future benefit of redistribution is zero for all agents.

The importance of mobility for political preferences has received less attention than cross-sectional inequality. But the simple model presented here shows that mobility is also an important factor in the determination of political preferences. More importantly, if inequality and mobility are not independent, either across countries or across times, by focusing only on inequality we may reach inaccurate conclusions. Suppose, for example, that an increase in cross-sectional inequality, σ, is associated with a decrease in ρ, that is, with an increase in mobility. Therefore, we have two contrasting effects: the increase in σ leads to higher taxes, whereas the decrease in ρ leads to lower taxes.

This example may help to explain why, in certain episodes of increasing inequality, such as in the United States before the recent crisis, we do not see a significant increase in demand for redistribution. Perhaps the reason is that voters perceive higher mobility as coincident with greater inequality. Then, thanks to the perceived mobility, voters do not demand higher taxes and the economy continues to perform well even if income becomes more concentrated. However, if the performance of the economy changes and voters start to perceive lower mobility, they will start demanding more redistribution, which will further deteriorate the performance of the economy. This idea has been developed in Quadrini (1999) in a model that features two equilibria. The first equilibrium is characterized by high growth, high inequality, and low redistribution. The second equilibrium is characterized by low growth, low inequality, and high redistribution.[16] The idea that the prospect of upward mobility reduces the demand for redistribution has also been studied in Benabou and Ok (2001).

14.5.1.2 The Case of Risk Aversion

We now assume that the utility function is concave and takes the following form:

$$u\left(c_t - \frac{h_t^2}{2}\right) = \frac{\left(c_t - \frac{h_t^2}{2}\right)^{1-\nu}}{1-\nu},$$

where the parameter ν captures the curvature of the utility function.

Figure 14.6 plots the preferred tax rate as a function of current endowment η_t^i for different values of ν. As can be seen from the figure, the preferred tax rate is monotonically decreasing in current endowment, and therefore, the median voter theorem also applies in the case of risk-averse agents. Furthermore, we see that, for each endowment level η_t^i, the preferred tax rate increases with risk aversion.

Figure 14.7 plots the preferred tax rate as a function of current endowment for different degrees of mobility and risk aversion. The first panel is for the case of risk neutrality. In this case, we see that lower mobility (ρ changes from 0.5 to 0.9) increases the equilibrium tax rate, that is, the tax rate preferred by the median voter, which in the figure is identified by the vertical line. These are the properties we have shown analytically in the previous subsection. However, when agents are risk averse, lower mobility reduces

[16] The increase in inequality in the United States is not a recent phenomenon. However, voters and politicians started to focus more on this issue after the recent crisis. During the good times in which financial markets were expanding, low-income households had access to credit, allowing them to own houses. For many this appeared as a new opportunity (mobility). However, with the crisis and the credit market and the freeze, these opportunities dried up and many households lost faith in the possibility of improving their current position (mobility). Not surprisingly, they turned to the government for help and asked for more populist policies.

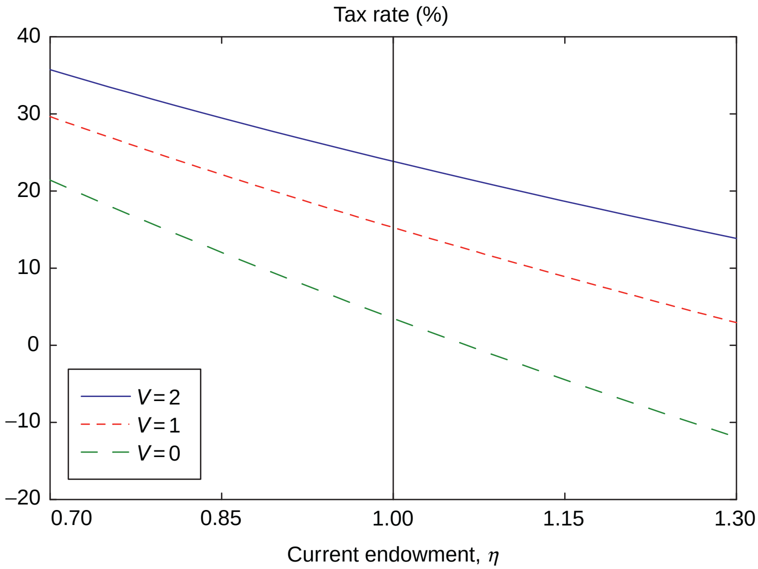


Figure 14.6 Preferred tax rates for different degrees of risk aversion. The inequality and mobility parameters are $\sigma = 0.5$ and $\rho = 0.5$. The vertical line denotes the median endowment.

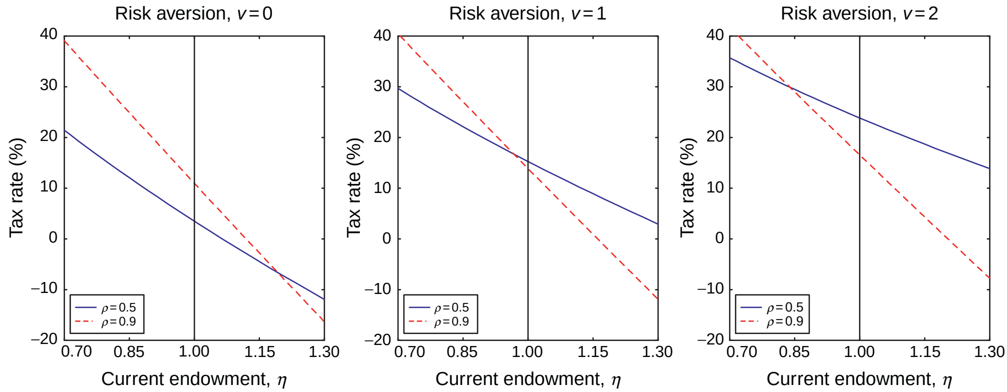


Figure 14.7 Preferred tax rates for different degrees of mobility, $\rho \in \{0.5, 0.9\}$, and risk aversion, $v \in \{0, 1, 2\}$. The inequality parameter is $\sigma = 0.5$. The vertical line denotes the median endowment.

the equilibrium tax rate. The reason is that, conditional on the current endowment, lower mobility means that agents face lower risk. In fact, with $\rho = 1$, next period endowment is equal to current endowment. Thus, there is less demand for insurance.

This example shows that mobility affects equilibrium policies through two mechanisms. The first mechanism works through the impact of mobility on the redistributive gains from next period taxes. When mobility is low, the expected redistributive gains are high. These gains vanish if mobility is perfect, that is, $\rho = 0$. The second mechanism works through the impact of mobility on individual risk. Given the current endowment, higher mobility (lower ρ) increases the conditional volatility of next period endowment and the median agent faces higher risk. Thus, greater redistribution is preferred if preferences are concave. This second mechanism is irrelevant when agents are risk neutral but becomes important when agents are risk averse. For a sufficiently high degree of risk aversion, the second mechanism dominates and the equilibrium tax rate declines with lower mobility.

Corbae et al. (2009) and Bachmann and Bai (2013) are two papers that study infinite horizon political economy models with income taxes and uninsurable idiosyncratic risks. Thus, these two papers are potentially capable of capturing the mechanisms described in this section.

14.5.2 More on the Political Economy Channel

Some theories formulate channels through which redistributive taxes have a beneficial effect on the macroeconomy in the presence of financial constraints. For example, in the Schumpeterian view where entrepreneurship is central to economic growth, financial constraints and the absence of insurance markets make entrepreneurial investment suboptimal. Under these conditions, redistribution may provide extra resources to constrained entrepreneurs and could facilitate more investments in growth-enhancing activities. At the same time, a redistributive system provides an implicit mechanism for consumption smoothing (a person pays high taxes when he or she earns high profits but receives payments in case of losses). Therefore, it provides insurance. Thus, if entrepreneurs are risk averse, redistribution could encourage investment.

A similar mechanism applies to the investment in education or human capital. If education is important for economic growth, and parents choose suboptimal levels of education because of financial constraints, then government transfers may allow for greater investment and growth. A more direct mechanism could work through the financing of public education, as in Glomm and Ravikumar (1992).

Political economy forces are also important for the choice of government borrowing. Azzimonti et al. (2014) propose a theory of public debt where greater income inequality could increase the incentive of the government to borrow more if the higher inequality is associated with greater individual risk. This is because higher risk increases the demand for safe assets, which are undersupplied when markets are incomplete. If financial markets are integrated, the increase in inequality (risk) in a few countries could induce a worldwide increase in public debt. In this way, the paper proposes one of the possible mechanisms for explaining the rising public debt observed in most of the industrialized countries since the early 1980s.

We close this section by mentioning that, although a large branch of the political economy literature has been developed on the assumption that voters are self-motivated and agree on their views of the world, so that their assessment of a policy is based on how much they benefit, some authors have proposed alternative frameworks. Especially interesting is Piketty (1995). This study develops a model in which agents prefer different policies not because they are selfish but because they have different beliefs. All voters care about social welfare, but some believe that luck is more important in generating income, whereas others believe that effort is more important. These beliefs evolve over time based on personal experience, but they never converge. Thus, at any point in time, preferences are heterogeneous. Although not explicitly explored in the original article, it is possible to introduce factors that could change the distribution of beliefs and with them the properties of the macro economy. This could be an interesting direction for future research.

14.6. CONCLUSION

In this chapter we have discussed a variety of topics that lie in the somewhat fuzzy intersection of income distribution and macroeconomics. The choice of topics and approaches has surely been idiosyncratic, reflecting our tastes, interests, and expertise, and we have left out many topics from behavioral and nonoptimizing models to issues in development, to the analysis of the impact of the rise of inequality on the U.S. economy. We have also touched other aspects only superficially, such as the role of globalization on the economy. In addition, we have only looked marginally at the implications of income inequality for consumption inequality or even for inequality in the duration of life,[17] which is in the end what really matters to determine the welfare costs of inequality.

We are very aware that a very different chapter covering the same could be written by other authors (in fact, the next chapter includes an example of this by providing some very different ideas of macro modeling of the wealth distribution). But we hope that this chapter has provided an idea of how macroeconomics is explicitly incorporating the analysis of inequality to improve our understanding of the dynamics of the aggregate economy, and also of how the discipline that macroeconomics brings to the table—that all pieces have to be mutually consistent and that dynamics is at the core of economics—shapes the way we think about income and wealth inequality.

ACKNOWLEDGMENTS

We are grateful for the comments from many people over the years. Many people contributed direct comments about and calculations for this paper. They include the editors for this handbook, but also Makoto Nakajima, Mariacristina De Nardi, Josep Pijoan-Mas, and Thomas Piketty. Others, such as David Wiczer and Moritz Kuhn, made even more direct contributions. We received research assistance from Kai Ding and Gero Dolfus, comments by Sergio Salgado and Annaliina Soikkanen, and editorial help from Joan Gieseke.

[17] Pijoan-Mas and Ríos-Rull (2014) argue that the welfare cost of inequality in life spans dwarfs that of inequality in consumption.

We thank all of them. Ríos-Rull thanks the National Science Foundation for grant SES-1156228. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

APPENDIX A. DERIVATION OF THE INEQUALITY INDEX

In each period, there are different cohorts of workers who have been employed for j periods. They also differ in terms of initial human capital h_0^k at birth. Because workers die with probability λ, the fraction of workers in the j cohort (composed of workers who survive for j periods) is equal to $\sum_k x_0^k(1-\lambda)^j$. Denote by h_j^k the human capital of a worker born with initial human capital h_0^k of age $j+1$. Because human capital grows at the gross rate $g(y,\varepsilon)$, we have that $h_j^k = h_0^k \prod_{t=1}^{j} g(y,\varepsilon_t)$. Of course, this differs across workers of the same cohort because the growth rate is stochastic. The average human capital is then computed as

$$\overline{h} = \sum_k x_0^k \sum_{j=0}^{\infty} (1-\lambda)^j \mathbb{E} h_j^k, \tag{A1}$$

where $\mathbb{E}$ averages the human capital of all agents in the $j-k$ cohort. Because growth rates are serially independent, we have that $\mathbb{E}h_j^k = h_0^k \mathbb{E}g(y,\varepsilon)^j$. Substituting in the above expression and solving we get

$$\overline{h} = \frac{\overline{h}_0}{1-(1-\lambda)\mathbb{E}g(y,\varepsilon)},$$

where $\overline{h}_0 = \sum_k x_0^k h_0^k$ is the aggregate human capital of newborn agents.

We now turn to the variance, which is calculated as

$$\mathrm{Var}(h) = \sum_k x_0^k \sum_{j=0}^{\infty} (1-\lambda)^j \mathbb{E}\left(h_j^k - \overline{h}\right)^2.$$

This can be rewritten as

$$\mathrm{Var}(h) = \sum_k x_0^k \sum_{j=0}^{\infty} (1-\lambda)^j \left[\mathbb{E}\left(h_j^k\right)^2 - \overline{h}^2\right].$$

This can be further rewritten as

$$\mathrm{Var}(h) = \sum_k x_0^k \left(h_0^k\right)^2 \sum_{j=0}^{\infty} (1-\lambda)^j \mathbb{E}\left(\frac{h_j^k}{h_0^k}\right)^2 - \overline{h}^2.$$

The term h_j^k/h_0^k is independent of the initial human capital h_0^k. Taking into account the serial independence of the growth rates, we have that $\mathbb{E}\left(h_j^k/h_0^k\right)^2 = \left[\mathbb{E}g(\gamma,\varepsilon)^2\right]^j$. Substituting and solving, we have

$$\mathrm{Var}(h) = \left(\sum_k x_0^k \left(h_0^k\right)^2\right)\left(\frac{1}{1-(1-\lambda)\mathbb{E}g(\gamma,\varepsilon)^2}\right) - \overline{h}^2.$$

To compute the inequality index, we simply divide the variance by $\overline{h}^2$, where $\overline{h}$ is given by Equation (A1). This returns the inequality index (14.27).

APPENDIX B. WAGE EQUATION WITH ENDOGENOUS DEBT

Consider the value of a filled vacancy defined in Equation (14.51). Using the binding enforcement $B' = \varphi(1-\eta)\mathbb{E}S'(B')$ to eliminate B', this value becomes

$$Q = (1+\varphi)\beta(1-\eta)\mathbb{E}S'(B').$$

Notice that at this stage we are imposing $b = B$ and $b' = B'$, which hold in a symmetric equilibrium.

Next we use the free entry condition $V = qQ - \kappa = 0$. Eliminating Q using the above expression and solving for the expected value of the surplus, we obtain

$$\mathbb{E}S'(B') = \frac{\kappa}{q(1+\varphi)\beta(1-\eta)}. \tag{B1}$$

Substituting into the definition of the surplus—Equation (14.50)—and taking into account that $b' = \varphi(1-\eta)\mathbb{E}S'(B')$, we get

$$S(B) = z - \overline{u} - B + \frac{[1-\lambda-p\eta+\varphi(1-\lambda)(1-\eta)]\kappa}{q(1+\varphi)(1-\eta)}. \tag{B2}$$

Now consider the net value for a worker,

$$W(B) - U = w - \overline{u} + \eta(1-\lambda-p)\beta\mathbb{E}S'(B').$$

Substituting $W(B) - U = \eta S(B)$ in the left-hand side and eliminating $\mathbb{E}S'(B')$ in the right-hand side using Equation (B1), we obtain

$$\eta S(B) = w - \overline{u} + \frac{\eta(1-\lambda-p)\kappa}{q(1+\varphi)(1-\eta)}. \tag{B3}$$

Finally, combining Equations (B2) and (B3) and solving for the wage, we get

$$w = (1-\eta)\overline{u} + \eta(z-b) + \frac{\eta[p+(1-\lambda)\varphi]\kappa}{q(1+\varphi)},$$

which is the expression reported in (14.54).

REFERENCES

Abowd, J.M., Card, D., 1987. Intertemporal labor supply and long-term employment contracts. Am. Econ. Rev. 77 (1), 50–68.

Acemoglu, D., 1998. Why do new technologies complement skills? Directed technical change and wage inequality. Q. J. Econ. 113 (4), 1055–1089.

Aghion, P., Bolton, P., 1997. A theory of trickle-down growth and development. Rev. Econ. Stud. 64 (2), 151–172.

Aguiar, M.A., Bils, M., 2011. Has consumption inequality mirrored income inequality? Working Paper 16807, National Bureau of Economic Research.

Aiyagari, S.R., 1994. Uninsured idiosyncratic risk and aggregate saving. Q. J. Econ. 109 (3), 659–684.

Alesina, A., Rodrik, D., 1994. Distributive politics and economic growth. Q. J. Econ. 109 (2), 465–490.

Alvaredo, F., 2011. A note on the relationship between top income shares and the Gini coefficient. Econ. Lett. 110 (3), 274–277.

Angeletos, G.-M., 2007. Uninsured idiosyncratic investment risk and aggregate saving. Rev. Econ. Dyn. 10 (1), 1–30.

Atkinson, A.B., Piketty, T., Saez, E., 2011. Top incomes in the long run of history. J. Econ. Lit. 49 (1), 3–71.

Attanasio, O., Battistin, E., Ichimura, H., 2004. What really happened to consumption inequality in the U.S.? Working Paper 10338, National Bureau of Economic Research.

Auerbach, A.J., Kotlikoff, L.J., 1987. Dynamic Fiscal Policy. Cambridge University Press, Cambridge.

Azzimonti, M., de Francisco, E., Quadrini, V., 2014. Financial globalization, inequality, and the rising public debt. Am. Econ. Rev. 104 (8), 2267–2302.

Bachmann, R., Bai, J.H., 2013. Politico-economic inequality and the comovement of government purchases. Rev. Econ. Dyn. 16 (4), 565–580.

Banerjee, A., Newman, A., 1993. Occupational choice and the process of development. J. Polit. Econ. 101 (2), 274–298.

Bell, B., Van Reenen, J., 2010. Bankers' pay and extreme wage inequality in the UK. Economic Performance Special Papers CEPSP21, Centre for Economic Performance, London School of Economics and Political Science.

Benabou, R., Ok, E.A., 2001. Social mobility and the demand for redistribution: the Poum hypothesis. Q. J. Econ. 116 (2), 447–487.

Benhabib, J., Rustichini, A., 1996. Social conflict and growth. J. Econ. Growth 1 (1), 125–142.

Benhabib, J., Bisin, A., Zhu, S., 2011. The distribution of wealth and fiscal policy in economies with finitely lived agents. Econometrica 79 (1), 123–157.

Ben-Porath, Y., 1967. The production of human capital and the life cycle of earnings. J. Polit. Econ. 75 (4), 352–365.

Bernanke, B., Gertler, M., 1989. Agency costs, net worth, and business fluctuations. Am. Econ. Rev. 79 (1), 14–31.

Bernanke, B.S., Gertler, M., Gilchrist, S., 1999. The financial accelerator in a quantitative business-cycle framework. In: Taylor, J.B., Woodford, M. (Eds.), Handbook of Macroeconomics, vol. 1. Elsevier, Amsterdam, pp. 1341–1393 (Chapter 21).

Bertola, G., 2000. Macroeconomics of distribution and growth. In: Atkinson, A., Bourguignon, F. (Eds.), Handbook of Income Distribution, vol. 1. Elsevier, Amsterdam, pp. 477–540 (Chapter 9).

Bewley, T., 1977. The permanent income hypothesis: a theoretical formulation. J. Econ. Theory 16 (2), 252–292.

Blanchard, O.J., 1985. Debt, deficits, and finite horizons. J. Polit. Econ. 93 (2), 223–247.

Budría, S., Díaz-Gimenez, J., Quadrini, V., Ríos-Rull, J.-V., 2001. New facts on the U.S. distribution of earnings, income and wealth. Unpublished Manuscript.

Buera, F.J., 2009. A dynamic model of entrepreneurship with borrowing constraints: theory and evidence. Ann. Finance 5 (3), 443–464.

Caballero, R.J., Farhi, E., Gourinchas, P.-O., 2008. An equilibrium model of "global imbalances" and low interest rates. Am. Econ. Rev. 98 (1), 358–393.

Cagetti, M., De Nardi, M., 2006. Entrepreneurship, frictions, and wealth. J. Polit. Econ. 114 (5), 835–869.

Cagetti, M., De Nardi, M., 2009. Estate taxation, entrepreneurship, and wealth. Am. Econ. Rev. 99 (1), 85–111.
Carlstrom, C.T., Fuerst, T.S., 1995. Interest rate rules vs. money growth rules: a welfare comparison in a cash-in-advance economy. J. Monet. Econ. 36 (2), 247–267.
Carrillo-Tudela, C., Visschers, L., 2013. Unemployment and endogenous reallocation over the business cycle. IZA Discussion Paper 7124, Institute for the Study of Labor (IZA).
Carroll, C.D., 1997. Buffer-stock saving and the life-cycle/permanent income hypothesis. Q. J. Econ. 112 (1), 1–55.
Caselli, F., 1999. Technological revolutions. Am. Econ. Rev. 89 (1), 78–102.
Caselli, F., Ventura, J., 2000. A representative consumer theory of distribution. Am. Econ. Rev. 90 (4), 909–926.
Castañeda, A., Díaz-Giménez, J., Ríos-Rull, J.-V., 1998. Exploring the income distribution business-cycle dynamics. J. Monet. Econ. 42 (1), 93–130.
Castañeda, A., Díaz-Giménez, J., Ríos-Rull, J.-V., 2003. Accounting for the U.S. earnings and wealth inequality. J. Polit. Econ. 111 (4), 818–857.
Chatterjee, S., 1994. Transitional dynamics and the distribution of wealth in a neoclassical growth model. J. Public Econ. 54 (1), 97–119.
Chatterjee, S., Corbae, D., Ríos-Rull, J.-V., 2004. A Competitive Theory of Credit Scoring. University of Pennsylvania, CAERP.
Chatterjee, S., Corbae, D., Nakajima, M., Ríos-Rull, J.-V., 2007. A quantitative theory of unsecured consumer credit with risk of default. Econometrica 75 (6), 1525–1589.
Chatterjee, S., Corbae, D., Ríos-Rull, J.-V., 2008. A finite-life private-information theory of unsecured consumer debt. J. Econ. Theory 142 (1), 149–177.
Cooley, T., Marimon, R., Quadrini, V., 2012. Risky investments with limited commitment. Working Paper 19594, National Bureau of Economic Research.
Corbae, D., D'Erasmo, P., Kuruscu, B., 2009. Politico-economic consequences of rising wage inequality. J. Monet. Econ. 56 (1), 43–61.
Costain, J.S., Reiter, M., 2008. Business cycles, unemployment insurance, and the calibration of matching models. J. Econ. Dyn. Control. 32 (4), 1120–1155.
Cummins, J.G., Violante, G.L., 2002. Investment-specific technical change in the United States (1947–2000): measurement and macroeconomic consequences. Rev. Econ. Dyn. 5 (2), 243–284.
De Nardi, M., 2004. Wealth inequality and intergenerational links. Rev. Econ. Stud. 71 (3), 743–768.
DeBacker, J., Heim, B., Panousi, V., Vidangos, I., 2011. Rising inequality: transitory or permanent? New evidence from a panel of U.S. household income 1987–2006. Finance and Economics Discussion Series 2011-60, Board of Governors of the Federal Reserve System (U.S.).
Díaz, A., Pijoan-Mas, J., Ríos-Rull, J.-V., 2003. Precautionary savings and wealth distribution under habit formation preferences. J. Monet. Econ. 50 (6), 1257–1291.
Díaz-Giménez, J., Pijoan-Mas, J., 2011. Flat tax reforms: investment expensing and progressivity. Discussion Paper No. 8238, Centre for Economic Policy Research.
Díaz-Giménez, J., Glover, A., Ríos-Rull, J.-V., 2011. Facts on the distributions of earnings, income, and wealth in the United States: 2007 update. Q. Rev. (Federal Reserve Bank of Minneapolis) 34 (1), 2–31.
Duras, J., 2013. Amplification of Shocks in a Model with Labor and Goods Market Search. University of Minnesota.
Fernandez-Villaverde, J., Krueger, D., 2011. Consumption and saving over the life cycle: how important are consumer durables? Macroecon. Dyn. 15 (5), 725–770.
Gabaix, X., Landier, A., 2008. Why has CEO pay increased so much? Q. J. Econ. 123 (1), 49–100.
Galor, O., Moav, O., 2000. Ability-biased technological transition, wage inequality, and economic growth. Q. J. Econ. 115 (2), 469–497.
Galor, O., Zeira, J., 1993. Income distribution and macroeconomics. Rev. Econ. Stud. 60 (1), 6035–6052.
Garicano, L., Rossi-Hansberg, E., 2004. Inequality and the organization of knowledge. Am. Econ. Rev. 94 (2), 197–202.
Gertler, M., Trigari, A., 2009. Unemployment fluctuations with staggered Nash wage bargaining. J. Polit. Econ. 117 (1), 38–86.

Glomm, G., Ravikumar, B., 1992. Public versus private investment in human capital: endogenous growth and income inequality. J. Polit. Econ. 100 (4), 818–834.
Glover, A., Heathcote, J., Krueger, D., Ríos-Rull, J.-V., 2011. Inter-generational redistribution in the Great Recession. Working Paper 16924, National Bureau of Economic Research.
Goldin, C., Katz, L.F., 2009. The Race Between Education and Technology. Belknap Press, Cambridge, MA.
Gordon, R.J., 1990. The measurement of durable goods prices. National Bureau of Economic Research Monograph Series, University of Chicago Press, Chicago, IL.
Gourinchas, P.-O., Parker, J.A., 2002. Consumption over the life cycle. Econometrica 70 (1), 47–89.
Greenwood, J., Jovanovic, B., 1990. Financial development, growth, and the distribution of income. J. Polit. Econ. 98 (5), 1076–1107.
Greenwood, J., Jovanovic, B., 1999. The information-technology revolution and the stock market. Am. Econ. Rev. Pap. Proc. 89 (2), 116–122.
Greenwood, J., Yorukoglu, M., 1997. Carnegie-Rochester conference series on public policy. J. Monet. Econ. 46 (2), 49–95.
Greenwood, J., Sanchez, J.M., Wang, C., 2010. Financing development: the role of information costs. Am. Econ. Rev. 100 (4), 1875–1891.
Guvenen, F., 2009. An empirical investigation of labor income processes. Rev. Econ. Dyn. 12 (1), 58–79.
Guvenen, F., Kuruşçu, B., 2010. A quantitative analysis of the evolution of the U.S. wage distribution, 1970–2000. In: Acemoglu, D., Rogoff, K., Woodford, M. (Eds.), NBER Macroeconomics Annual, vol. 24, pp. 227–276.
Guvenen, F., Kuruşçu, B., 2012. Understanding the evolution of the U.S. wage distribution: a theoretical analysis. J. Eur. Econ. Assoc. 10 (3), 489–517.
Guvenen, F., Kuruşçu, B., Ozkan, S., 2009. Taxation of human capital and wage inequality: a cross-country analysis. Federal Reserve Bank of Minneapolis, Research Department Staff Report 438.
Guvenen, F., Ozkan, S., Song, J., 2012. The nature of countercyclical income risk. Working Paper 18035, National Bureau of Economic Research.
Hagedorn, M., Manovskii, I., 2008. The cyclical behavior of equilibrium unemployment and vacancies revisited. Am. Econ. Rev. 98 (4), 1692–1706.
Hansen, G.D., Imrohoroglu, S., 2008. Consumption over the life cycle: the role of annuities. Rev. Econ. Dyn. 11 (3), 566–583.
Heaton, J., Lucas, D.J., 1996. Evaluating the effects of incomplete markets on risk sharing and asset pricing. J. Polit. Econ. 104 (3), 443–487.
Heckman, J., Lochner, L., Taber, C., 1998. Explaining rising wage inequality: explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents. Rev. Econ. Dyn. 1 (1), 1–58.
Hobijn, B., Jovanovic, B., 2001. The information-technology revolution and the stock market: evidence. Am. Econ. Rev. 91 (5), 1203–1220.
Hornstein, A., Krusell, P., Violante, G., 2011. Frictional wage dispersion in search models: a quantitative assessment. Am. Econ. Rev. 101 (7), 2873–2898.
Huggett, M., 1993. The risk-free rate in heterogeneous-agent, incomplete-insurance economies. J. Econ. Dyn. Control. 17 (5), 953–969.
Huggett, M., 1996. Wealth distribution in life-cycle economies. J. Monet. Econ. 38 (3), 469–494.
Huggett, M., Ventura, G., Yaron, A., 2011. Sources of lifetime inequality. Am. Econ. Rev. 101 (7), 2923–2954.
İmrohoroğlu, A., 1989. Cost of business cycles with indivisibilities and liquidity constraints. J. Polit. Econ. 97 (6), 1364–1383.
Kambourov, G., Manovskii, I., 2009a. Occupational mobility and wage inequality. Rev. Econ. Stud. 76 (2), 731–759.
Kambourov, G., Manovskii, I., 2009b. Occupational specificity of human capital. Int. Econ. Rev. 50 (1), 63–115.
Karabarbounis, L., Neiman, B., 2014. The global decline of the labor share. Q. J. Econ. 129 (1), 61–103.
Katz, L.F., Murphy, K.M., 1992. Changes in relative wages, 1963–1987: supply and demand factors. Q. J. Econ. 107 (1), 35–78.

Kehoe, T., Levine, D., 1993. Debt-constrained asset markets. Rev. Econ. Stud. 60 (4), 865–888.
Kiyotaki, N., Moore, J., 1997. Credit cycles. J. Polit. Econ. 105 (2), 211–248.
Krueger, D., Kubler, F., 2006. Pareto-improving social security reform when financial markets are incomplete!? Am. Econ. Rev. 96 (3), 737–755.
Krueger, D., Perri, F., 2006. Does income inequality lead to consumption inequality? Evidence and theory. Rev. Econ. Stud. 73 (1), 163–193.
Krueger, D., Perri, F., Pistaferri, L., Violante, G.L., 2010. Cross-sectional facts for macroeconomists. Rev. Econ. Dyn. 13 (1), 1–14.
Krusell, P., Ríos-Rull, J.-V., 1996. Vested interests in a positive theory of stagnation and growth. Rev. Econ. Stud. 63 (215), 301–331.
Krusell, P., Smith Jr., A.A., 1998. Income and wealth heterogeneity in the macroeconomy. J. Polit. Econ. 106 (5), 867–896.
Krusell, P., Quadrini, V., Ríos-Rull, J.-V., 1997. Politico-economic equilibrium and economic growth. J. Econ. Dyn. Control. 21 (1), 243–272.
Krusell, P., Ohanian, L.E., Ríos-Rull, J.-V., Violante, G.L., 2000. Capital-skill complementarity and inequality: a macroeconomic analysis. Econometrica 68 (5), 1029–1054.
Kuhn, M., 2014. Trends in income and wealth inequality. University of Bonn, Working Paper.
Kumhof, M., Rancière, R., 2010. Inequality, leverage and crises. IMF Working Paper WP/10/268.
Livshits, I., MacGee, J., Tertilt, M., 2007. Consumer bankruptcy: a fresh start. Am. Econ. Rev. 97 (1), 402–418.
Manuelli, R.E., Seshadri, A., 2010. Human capital and the wealth of nations. Unpublished Manuscript, University of Wisconsin-Madison.
Meh, C.A., Quadrini, V., 2006. Endogenous market incompleteness with investment risks. J. Econ. Dyn. Control 30 (11), 2143–2165.
Meltzer, A.H., Richard, S.F., 1981. A rational theory of the size of government. J. Polit. Econ. 89 (5), 914–927.
Mendoza, E.G., Quadrini, V., Ríos-Rull, J.-V., 2007. Financial integration, financial deepness and global imbalances. Working Paper 12909, National Bureau of Economic Research.
Milgrom, P., Roberts, J., 1990. The economics of modern manufacturing: technology, strategy, and organization. Am. Econ. Rev. 80 (3), 511–528.
Modigliani, F., Miller, M.H., 1958. The cost of capital, corporate finance and the theory of investment. Am. Econ. Rev. 48 (3), 261–279.
Monacelli, T., Quadrini, V., Trigari, A., 2011. Financial markets and unemployment. Working Paper 17389, National Bureau of Economic Research.
Murphy, K.M., Shleifer, A., Vishny, R., 1989. Income distribution, market size, and industrialization. Q. J. Econ. 104 (3), 537–564.
Papageorgiou, T., 2009. Learning your comparative advantages. 2009 Meeting Papers 1150, Society for Economic Dynamics.
Persson, T., Tabellini, G., 1994. Is inequality harmful for growth? Am. Econ. Rev. 84 (3), 600–621.
Pijoan-Mas, J., Ríos-Rull, J.-V., 2014. The Welfare Cost of Inequality in Life Expectancies. University of Minnesota.
Piketty, T., 1995. Social mobility and redistributive politics. Q. J. Econ. 110 (3), 551–584.
Piketty, T., 2014. Capital in the Twenty-First Century. Harvard University Press, Cambridge, MA. ISBN 9780674369559.
Piketty, T., Saez, E., 2003. Income inequality in the United States, 1913–1998. Q. J. Econ. 118 (1), 1–39. Longer updated version published in: Atkinson, A.B., Piketty, T. (Eds.), Top Incomes over the Twentieth Century: A Contrast between European and English-Speaking Countries. Oxford University Press, Oxford, 2007.
Pissarides, C.A., 1987. Search, wage bargains and cycles. Rev. Econ. Stud. 54 (3), 473–483.
Pissarides, C.A., 1990. Equilibrium Unemployment Theory. Basil Blackwell, Oxford.
Quadrini, V., 1999. Growth, learning and redistributive policies. J. Public Econ. 74 (2), 263–297.
Quadrini, V., 2000. Entrepreneurship, saving, and social mobility. Rev. Econ. Dyn. 3 (1), 1–40.
Quadrini, V., 2011. Financial frictions in macroeconomic fluctuations. Econ. Q. 97 (3Q), 209–254.

Ríos-Rull, J.-V., 1994. On the quantitative importance of market completeness. J. Monet. Econ. 34 (3), 463–496.
Ríos-Rull, J.-V., 1996. Life-cycle economies and aggregate fluctuations. Rev. Econ. Stud. 63 (3), 465–489.
Ríos-Rull, J.-V., Santaeulàlia-Llopis, R., 2010. Redistributive shocks and productivity shocks. J. Monet. Econ. 57 (8), 931–948.
Rosen, S., 1981. The economics of superstars. Am. Econ. Rev. 71 (5), 845–858.
Saez, E., Zucman, G., 2014. The distribution of US wealth, capital income and returns since 1913. Unpublished Manuscript, University of California, Berkeley.
Shimer, R., 2005. The cyclical behavior of equilibrium unemployment and vacancies. Am. Econ. Rev. 95 (1), 25–49.
Solow, R., 1987. We'd Better Watch Out. New York Times Book Review, p. 36, July.
Stiglitz, J.E., 1969. Distribution of income and wealth among individuals. Econometrica 37 (3), 382–397.
Storesletten, K., Telmer, C.I., Yaron, A., 2001. How important are idiosyncratic shocks? Evidence from labor supply. Am. Econ. Rev. Pap. Proc. 91 (2), 413–417.
Terajima, Y., 2006. Education and self-employment: changes in earnings and wealth inequality. Bank of Canada Working Paper 2006-40.
Townsend, R.M., 1979. Optimal contracts and competitive markets with costly state verification. J. Econ. Theory 21 (2), 265–293.
von Weizsäcker, R.K., 1993. A Theory of Earnings Distribution. Cambridge University Press, Cambridge.
Wiczer, D., 2013. Long-term unemployment: attached and mismatched? Unpublished Manuscript, University of Minnesota.
Yaari, M.E., 1965. Uncertain lifetime, life insurance and the theory of the consumer. Rev. Econ. Stud. 32 (2), 137–150.
Yamaguchi, S., 2012. Tasks and heterogeneous human capital. J. Labor Econ. 30 (1), 1–53.

CHAPTER 15

Wealth and Inheritance in the Long Run

Thomas Piketty*, Gabriel Zucman†
*Paris School of Economics, Paris, France
†London School of Economics, London, UK

Contents

Handbook of Income Distribution, Volume 2B
ISSN 1574-0056, http://dx.doi.org/10.1016/B978-0-444-59429-7.00016-9

Abstract

This chapter offers an overview of the empirical and theoretical research on the long-run evolution of wealth and inheritance. Wealth–income ratios, inherited wealth, and wealth inequalities were high in the eighteenth to nineteenth centuries up until World War I, then sharply dropped during the twentieth century following World War shocks, and have been rising again in the late twentieth and early twenty-first centuries. We discuss the models that can account for these facts. We show that over a wide range of models, the long-run magnitude and concentration of wealth and inheritance are an increasing function of $\bar{r} - g$ where $\bar{r}$ is the net-of-tax rate of return on wealth and g is the economy's growth rate. This suggests that current trends toward rising wealth–income ratios and wealth inequality might continue during the twenty-first century, both because of the slowdown of population and productivity growth, and because of rising international competition to attract capital.

Keywords

Wealth, Inheritance, Distribution, Growth, Rate of return, Pareto coefficient

JEL Classification Codes

E10, D30, D31, D32

15.1. INTRODUCTION

Economists have long recognized that the magnitude and distribution of wealth play an important role in the distribution of income—both across factors of production (labor and capital) and across individuals. In this chapter, we ask three simple questions: (1) What do we know about historical patterns in the magnitude of wealth and inheritance relative to income? (2) How does the distribution of wealth vary in the long run and across countries? (3) And what are the models that can account for these facts?

In surveying the literature on these issues, we will focus the analysis on three interrelated ratios. The first is the aggregate wealth-to-income ratio, that is the ratio between marketable—nonhuman—wealth and national income. The second is the share of aggregate wealth held by the richest individuals, say the top 10% or top 1%. The last is the ratio between the stock of inherited wealth and aggregate wealth (or between the annual flow of bequests and national income). As we shall see, to properly analyze the concentration of wealth and its implications, it is critical to study top wealth shares jointly with the macroeconomic wealth–income and inheritance–wealth ratios. In so doing, this chapter attempts to build bridges between income distribution and macroeconomics.

The wealth-to-income ratio, top wealth shares, and the share of inheritance in the economy have all been the subject of considerable interest and controversy—but usually on the basis of limited data. For a long time, economics textbooks have presented the wealth–income ratio as stable over time—one of the Kaldor facts.[1] There is, however, no strong theoretical reason why it should be so: With a flexible production function, any ratio can be a steady state. And until recently, we lacked comprehensive national balance sheets with harmonized definitions for wealth that could be used to vindicate the constant-ratio thesis. Recent research shows that wealth–income ratios, as well as the share of capital in national income, are actually much less stable in the long run than what is commonly assumed.

Following the Kuznets curve hypothesis, first formulated in the 1950s, another common view among economists has been that income inequality—and possibly wealth inequality as well—should first rise and then decline with economic development, as a growing fraction of the population joins high-productivity sectors and benefits from industrial growth.[2] However, following the rise in inequality that has occurred in most developed countries since the 1970s–1980s, this optimistic view has become less popular.[3] As a consequence, most economists are now fairly skeptical about universal laws regarding the long-run evolution of inequality.

Last, regarding the inheritance share in total wealth accumulation, there seems to exist a general presumption that it should tend to decline over time. Although this is rarely formulated explicitly, one possible mechanism could be the rise of human capital (leading maybe to a rise of the labor share in income and saving), or the rise in life-cycle wealth accumulation (itself possibly due to the rise of life expectancy). Until recently, however, there was limited empirical evidence on the share of inherited wealth available to test these hypotheses. The 1980s saw a famous controversy between Modigliani (a life-cycle advocate, who argued that the share of inherited wealth was as little as 20–30% of U.S. aggregate wealth) and Kotlikoff–Summers (who instead

[1] See, e.g., Kaldor (1961) and Jones and Romer (2010).

[2] See Kuznets (1953).

[3] See Atkinson et al. (2011). See also Chapter 7 in Handbook of Income Distribution, volume 2A by Roine and Waldenstrom (2015).

argued that the inheritance share was as large as 80%, if not larger). Particularly confusing was the fact that both sides claimed to look at the same data, namely U.S. data from the 1960s–1970s.[4]

Because many of the key predictions about wealth and inheritance were formulated a long time ago—often in the 1950s–1960s, or sometime in the 1970s–1980s—and usually on the basis of a relatively small amount of long-run evidence, it is high time to take a fresh look at them again on the basis of the more reliable evidence now available.

We begin by reviewing in Section 15.2 what we know about the historical evolution of the wealth–income ratio β. In most countries, this ratio has been following a U-shaped pattern over the 1910–2010 period, with a large decline between the 1910s and the 1950s, and a gradual recovery since the 1950s. The pattern is particularly spectacular in Europe, where the aggregate wealth–income ratio was as large as 600–700% during the eighteenth, nineteenth, and early twentieth centuries, then dropped to as little as 200–300% in the mid-twentieth century. It is now back to about 500–600% in the early twenty-first century. These same orders of magnitude also seem to apply to Japan, though the historical data is less complete than for Europe. The U-shaped pattern also exists—but is less marked—in the United States.

In Section 15.3, we turn to the long-run changes in wealth concentration. We also find a U-shaped pattern over the past century, but the dynamics have been quite different in Europe and in the United States. In Europe, the recent increase in wealth inequality appears to be more limited than the rise of the aggregate wealth–income ratio, so that European wealth seems to be significantly less concentrated in the early twenty-first century than a century ago. The top 10% wealth share used to be as large as 90%, whereas it is around 60–70% today (which is already quite large—and in particular a lot larger than the concentration of labor income). In the United States, by contrast, wealth concentration appears to have almost returned to its early twentieth century level. Although Europe was substantially more unequal than the United States until World War I, the situation has reversed over the course of the twentieth century. Whether the gap between both economies will keep widening in the twenty-first century is an open issue.

In Section 15.4, we describe the existing evidence regarding the evolution of the share φ of inherited wealth in aggregate wealth. This is an area in which available historical series are scarce and a lot of data has yet to be collected. However existing evidence—coming mostly from France, Germany, the United Kingdom, and Sweden—suggests that the inheritance share has also followed a U-shaped pattern over the past century. Modigliani's estimates—with a large majority of wealth coming from life-cycle savings—might have been right for the immediate postwar period (though somewhat exaggerated). But Kotlikoff–Summers' estimates—with inheritance

[4] See Kotlikoff and Summers (1981, 1988) and Modigliani (1986, 1988). Modigliani's theory of life-cycle saving was first formulated in the 1950s–1960s; see the references given in Modigliani (1986).

accounting for a significant majority of wealth—appear to be closer to what we generally observe in the long run, both in the nineteenth, twentieth, and early twenty-first centuries. Here again, there could be some interesting differences between Europe and the United States (possibly running in the opposite direction than for wealth concentration). Unfortunately the fragility of available U.S. data makes it difficult to conclude at this stage.

We then discuss in Section 15.5 the theoretical mechanisms that can be used to account for the historical evidence and to analyze future prospects. Some of the evolutions documented in Sections 15.2–15.4 are due to shocks. In particular, the large U-shaped pattern of wealth–income and inheritance-income ratios observed over the 1910–2010 period is largely due to the wars (which hit Europe and Japan much more than the United States). Here the main theoretical lesson is simply that capital accumulation takes time, and that the world wars of the twentieth century have had a long-lasting impact on basic economic ratios. This, in a way, is not too surprising and follows from simple arithmetic. With a 10% saving rate and a fixed income, it takes 50 years to accumulate the equivalent of 5 years of income in capital stock. With income growth, the recovery process takes even more time.

The more interesting and difficult part of the story is to understand the forces that determine the new steady-state levels toward which each economy tends to converge once it has recovered from shocks. In Section 15.5, we show that over a wide range of models, the long-run magnitude and concentration of wealth and inheritance are a decreasing function of g and an increasing function of $\bar{r}$, where g is the economy's growth rate and $\bar{r}$ is the net-of-tax rate of return to wealth. That is, under plausible assumptions, our three interrelated sets of ratios—the wealth–income ratio, the concentration of wealth, and the share of inherited wealth—all tend to take higher steady-state values when the long-run growth rate is lower or when the net-of-tax rate of return is higher. In particular, a higher $\bar{r}-g$ tends to magnify steady-state wealth inequalities. We argue that these theoretical predictions are broadly consistent with both the time-series and the cross-country evidence. This also suggests that the current trends toward rising wealth–income ratios and wealth inequality might continue during the twenty-first century, both because of population and productivity growth slowdown, and because of rising international competition to attract capital.

Owing to data availability constraints, the historical evolutions analyzed in this chapter relate for the most part to today's rich countries (Europe, North America, and Japan). However, to the extent that the theoretical mechanisms unveiled by the experience of rich countries also apply elsewhere, the findings presented here are also of interest for today's emerging economies. In Section 15.5, we discuss the prospects for the global evolution of wealth–income ratios, wealth concentration, and the share of inherited wealth in the coming decades. Finally, Section 15.6 offers concluding comments and stresses the need for more research in this area.

15.2. THE LONG-RUN EVOLUTION OF WEALTH–INCOME RATIOS

15.2.1 Concepts, Data Sources, and Methods

15.2.1.1 Country Balance Sheets

Prior to World War I, there was a vibrant tradition of national wealth accounting: economists, statisticians, and social arithmeticians were much more interested in computing the stock of national wealth than the flows of national income and output. The first national balance sheets were established in the late seventeenth and early eighteenth centuries by Petty (1664) and King (1696) in the United Kingdom, and Boisguillebert (1695) and Vauban (1707) in France. National wealth estimates then became plentiful in the nineteenth and early twentieth century, with the work of Colquhoun (1815), Giffen (1889), and Bowley (1920) in the United Kingdom, de Foville (1893) and Colson (1903) in France, Helfferich (1913) in Germany, King (1915) in the United States, and dozens of other economists.

The focus on wealth, however, largely disappeared in the interwar. The shock of World War I, the Great Depression, and the coming of Keynesian economics led to attention being switched from stocks to flows, with balance sheets being neglected. The first systematic attempt to collect historical balance sheets is due to Goldsmith (1985, 1991). Building upon recent progress made in the measurement of wealth, and pushing forward Goldsmith's pioneering attempt, Piketty and Zucman (2014) construct aggregate wealth and income series for the top eight rich economies. Other recent papers that look at specific countries include Atkinson (2013) for the United Kingdom and Ohlsson et al. (2013) for Sweden. In this section, we rely on the data collected by Piketty and Zucman (2014)—and closely follow the discussion therein—to present the long-run evolution of wealth–income ratios in the main developed economies.

In determining what is to be counted as wealth, we follow the U.N. System of National Accounts (SNA). For the 1970–2010 period, the data come from official national accounts that comply with the latest international guidelines (SNA, 1993, 2008). For the previous periods, Piketty and Zucman (2014) draw on the vast national wealth accounting tradition to construct homogenous income and wealth series that use the same concepts and definitions as in the most recent official accounts. The historical data themselves were established by a large number of scholars and statistical administrations using a wide variety of sources, including land, housing and wealth censuses, financial surveys, corporate book accounts, and the like. Although historical balance sheets are far from perfect, their methods are well documented and they are usually internally consistent. It was also somewhat easier to estimate national wealth around 1900–1910 than it is today: the structure of property was simpler, with less financial intermediation and cross-border positions.[5]

[5] A detailed analysis of conceptual and methodological issues regarding wealth measurement, as well as extensive country-specific references on historical balance sheets, are provided by Piketty and Zucman (2014).

15.2.1.2 Concepts and Definitions: Wealth Versus Capital

We define private wealth, W_t, as the net wealth (assets minus liabilities) of households.[6] Following SNA guidelines, assets include all the nonfinancial assets—land, buildings, machines, etc.—and financial assets—including life insurance and pensions funds—over which ownership rights can be enforced and that provide economic benefits to their owners. Pay-as-you-go Social Security pension wealth is excluded, just as all other claims on future government expenditures and transfers (such as education expenses for one's children or health benefits). Durable goods owned by households, such as cars and furniture, are excluded as well.[7] As a general rule, all assets and liabilities are valued at their prevailing market prices. Corporations are included in private wealth through the market value of equities and corporate bonds. Unquoted shares are typically valued on the basis of observed market prices for comparable, publicly traded companies.

Similarly, public (or government) wealth, W_{gt}, is the net wealth of public administrations and government agencies. In available balance sheets, public nonfinancial assets such as administrative buildings, schools, and hospitals are valued by cumulating past investment flows and upgrading them using observed real estate prices.

Market-value national wealth, W_{nt}, is the sum of private and public wealth:

$$W_{nt} = W_t + W_{gt}$$

and national wealth can also be decomposed into domestic capital and net foreign assets:

$$W_{nt} = K_t + NFA_t$$

In turn, domestic capital K_t can be written as the sum of agricultural land, housing, and other domestic capital (including the market value of corporations, and the value of other nonfinancial assets held by the private and public sectors, net of their liabilities).

Regarding income, the definitions and notations are standard. Note that we always use net-of-depreciation income and output concepts. National income Y_t is the sum of net domestic output and net foreign income: $Y_t = Y_{dt} + r_t \cdot NFA_t$.[8] Domestic output can be thought of as coming from some aggregate production function that uses domestic capital and labor as inputs: $Y_{dt} = F(K_t, L_t)$.

[6] Private wealth also includes the assets and liabilities held by nonprofit institutions serving households (NPISH). The main reason for doing so is that the frontier between individuals and private foundations is not always clear. In any case, the net wealth of NPISH is usually small, and always less than 10% of total net private wealth: currently it is about 1% in France, 3–4% in Japan, and 6–7% in the United States; see Piketty and Zucman (2014, Appendix Table A65). Note also that the household sector includes all unincorporated businesses.

[7] The value of durable goods appears to be relatively stable over time (about 30–50% of national income, i.e., 5–10% of net private wealth). See for instance Piketty and Zucman (2014, Appendix Table US.6f) for the long-run evolution of durable goods in the United States.

[8] National income also includes net foreign labor income and net foreign production taxes—both of which are usually negligible.

One might prefer to think about output as deriving from a two-sector production process (housing and nonhousing sectors), or more generally from n sectors. In the real world, the capital stock K_t comprises thousands of various assets valued at different prices (just like output Y_{dt} is defined as the sum of thousands of goods and services valued at different prices). We find it more natural, however, to start with a one-sector formulation. Since the same capital assets (i.e., buildings) are often used for housing and office space, it would be quite artificial to start by dividing capital and output into two parts. We will later on discuss the pros and cons of the one-sector model and the need to appeal to two-sector models and relative asset price movements to properly account for observed changes in the aggregate wealth–income ratio.

Another choice that needs to be discussed is the focus on market values for national wealth and capital. We see market values as a useful and well-defined starting point. But one might prefer to look at book values, for example, for short-run growth accounting exercises. Book values exceed market values when Tobin's Q is less than 1, and conversely when Tobin's Q is larger than 1. In the long run, however, the choice of book versus market value does not much affect the analysis (see Piketty and Zucman, 2014, for a detailed discussion).

We are interested in the evolution of the private wealth–national income ratio $\beta_t = W_t/Y_t$ and of the national wealth–national income ratio $\beta_{nt} = W_{nt}/Y_t$. In a closed economy, and more generally in an open economy with a zero net foreign position, the national wealth–national income ratio β_{nt} is the same as the domestic capital–output ratio $\beta_{kt} = K_t/Y_{dt}$.[9] If public wealth is equal to zero, then both ratios are also equal to the private wealth–national income ratio $\beta_t = \beta_{nt} = \beta_{kt}$. At the global level, the world wealth–income ratio is always equal to the world capital–output ratio.

15.2.2 The Very Long-Run: Britain and France, 1700–2010

Figures 15.1 and 15.2 present the very long-run evidence available for Britain and France regarding the national wealth–national income ratio β_{nt}. Net public wealth—either positive or negative—is usually a relatively small fraction of national wealth, so that the evolution of β_{nt} mostly reflects the evolution of the private wealth–national income ratio β_t (more on this below).[10]

[9] In principle, one can imagine a country with a zero net foreign asset position (so that $W_{nt} = K_t$) but nonzero net foreign income flows (so that $Y_t \neq Y_{dt}$). In this case the national wealth–national income ratio β_{nt} will slightly differ from the domestic capital–output ratio β_{kt}. In practice today, differences between Y_t and Y_{dt} are very small—national income Y_t is usually between 97% and 103% of domestic output Y_{dt} (see Piketty and Zucman, 2014, Appendix Figure A57). Net foreign asset positions are usually small as well, so that β_{kt} turns out to be usually close to β_{nt} in the 1970–2010 period (see Piketty and Zucman, 2014, Appendix Figure A67).

[10] For an historical account of the changing decomposition of national wealth into private and public wealth in Britain and France since the eighteenth century, see Piketty (2014, Chapter 3).

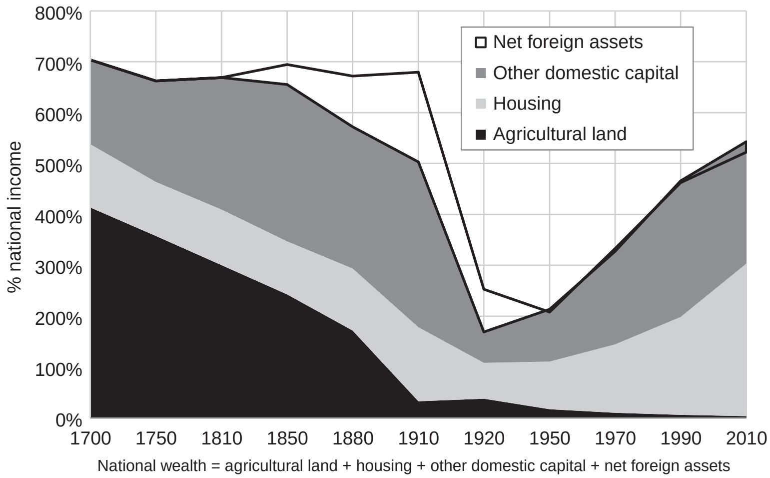


Figure 15.1 The changing level and nature of national wealth: United Kingdom 1700–2010.

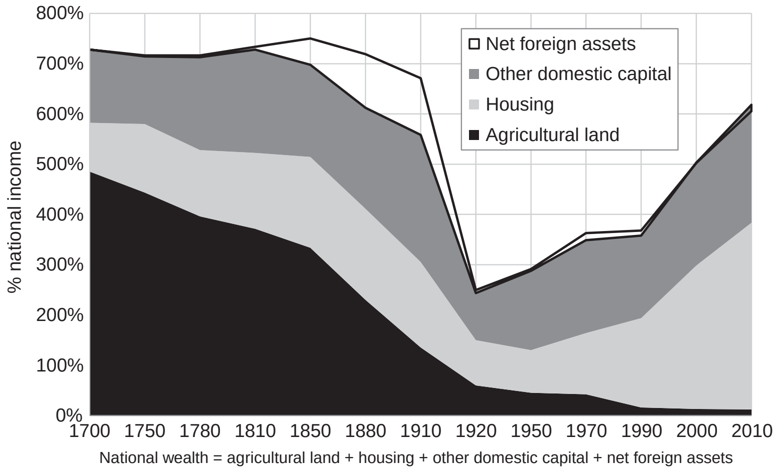


Figure 15.2 The changing level and nature of national wealth: France 1700–2010.

The evolutions are remarkably similar in the two countries. First, the wealth–income ratio has followed a spectacular U-shaped pattern. Aggregate wealth was worth about 6–7 years of national income during the eighteenth to nineteenth centuries on both sides of the channel, up until the eve of World War I. Raw data sources available for these two centuries are not sufficiently precise to make fine comparisons between the two countries or over time, but the orders of magnitude appear to be reliable and roughly stable (they come from a large number of independent estimates). Aggregate wealth then collapsed to as little as 2–3 years of national income in the aftermath of the two World Wars. Since the 1950s, there has been a gradual recovery in both countries. Aggregate wealth is back to about 5–6 years of national income in the 2000s to 2010s, just a bit below the pre-World War I level.

The other important finding that emerges from Figures 15.1 and 15.2 is that the composition of national wealth has changed in similar ways in both countries. Agricultural land, which made the majority of national capital in the eighteenth century, has been gradually replaced by real estate and other domestic capital, which is for the most part business capital (i.e., structures and equipment used by private firms). The nature of wealth has changed entirely reflecting a dramatic change in the structure of economic activity, and yet the total value of wealth is more or less the same as what it used to be before the Industrial Revolution.

Net foreign assets also made a large part of national capital in the late nineteenth century and on the eve of World War I: as much as 2 years of national capital in the case of Britain and over a year in the case of France. Net foreign-asset positions were brought back to zero in both countries following World War I and II shocks (including the loss of the colonial empires). In the late twentieth and early twenty-first centuries, net foreign positions are close to zero in both countries, just as in the eighteenth century. In the very long run, net foreign assets do not matter too much for the dynamics of the capital/income ratio in Britain or France. The main structural change is the replacement of agricultural land by housing and business capital.[11]

15.2.3 Old Europe Versus the New World

It is interesting to contrast the case of Old Europe—as illustrated by Britain and France—with that of the United States.

As Figure 15.3 shows, the aggregate value of wealth in the eighteenth to nineteenth centuries was markedly smaller in the New World than in Europe. At the time of the Declaration of Independence and in the early nineteenth century, national wealth in the United States was barely equal to 3–4 years of national income, about half that of Britain or France. Although available estimates are fragile, the order of magnitude again

[11] It is worth stressing that should we divide aggregate wealth by disposable household income (rather than national income), then today's ratios would be around 700–800% in Britain or France and would slightly surpass eighteenth to nineteenth century levels. This mechanically follows from the fact that disposable income was above 90% in the eighteenth to nineteenth centuries and is about 70–80% of disposable income in the late twentieth to early twenty-first century. The rising gap between disposable and household income reflects the rise of government-provided services, in particular in health and education. To the extent that these services are mostly useful (in their absence households would have to purchase them on the market), it is more justified for the purpose of historical and international comparisons to focus on ratios using national income as a denominator. For wealth–income ratios using disposable income as a denominator, see Piketty and Zucman (2014, Appendix, Figure A9).

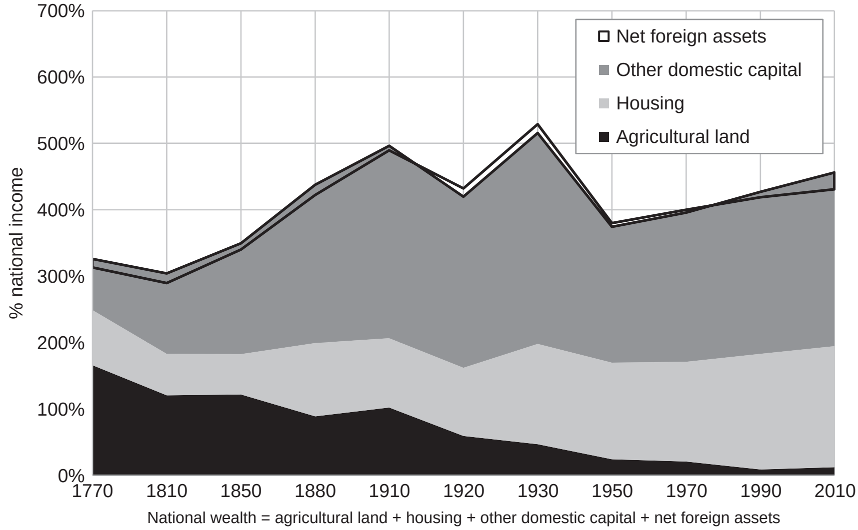


Figure 15.3 The changing level and nature of national wealth: United States 1770–2010.

appears to be robust. In Section 15.5, we will attempt to account for this interesting contrast. At this stage, we simply note that there are two obvious—and potentially complementary—factors that can play a role: first, there had been less time to save and accumulate wealth in the New World than in the Old World; second, there was so much land in the New World that it was almost worthless (its market value per acre was much less than in Europe).

The gap between the United States and Europe gradually reduces over the course of the nineteenth century, but still remains substantial. Around 1900–1910, national wealth is about 5 years of national income in the United States (see Figure 15.3) versus about 7 years in Britain and France. During the twentieth century, the U.S. wealth–income ratio also follows a U-shaped pattern, but less marked than in Europe. National wealth falls less sharply in the United States than in Europe following World War shocks, which seems rather intuitive. Interestingly, European wealth–income ratios have again surpassed U.S. ratios in the late twentieth and early twenty-first centuries.

This brief overview of wealth in the New World and Europe would be rather incomplete if we did not mention the issue of slavery. As one can see from Figure 15.4, the aggregate market value of slaves was fairly substantial in the United States until 1865: about 1–1.5 years of national income according to the best available historical sources. There were few slaves in Northern states, but in the South the value of the slave stock was so large that it approximately compensated—from the viewpoint of slave owners—the lower value of land as compared to the Old World (see Figure 15.5).

It is rather dubious, however, to include the market value of slaves into national capital. Slavery can be viewed as the most extreme form of debt: it should be counted as an

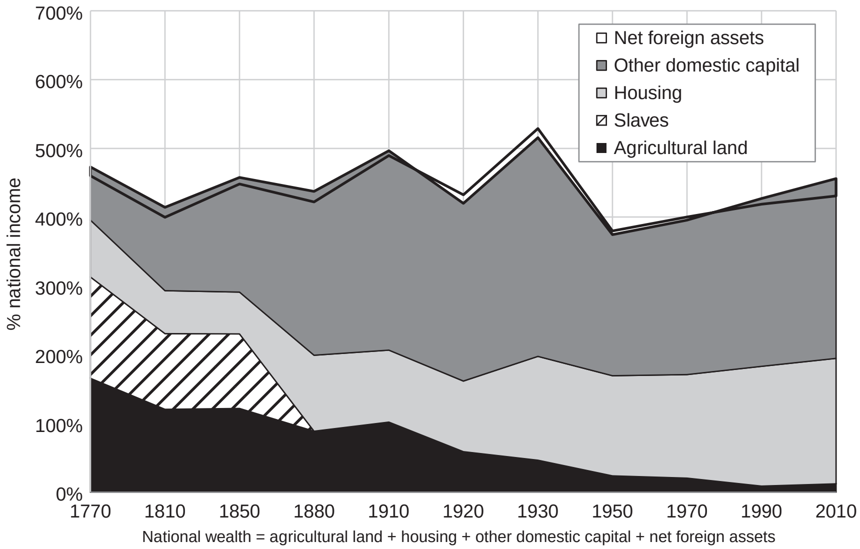


Figure 15.4 The changing level and nature of wealth: United States 1770–2010 (including slaves).

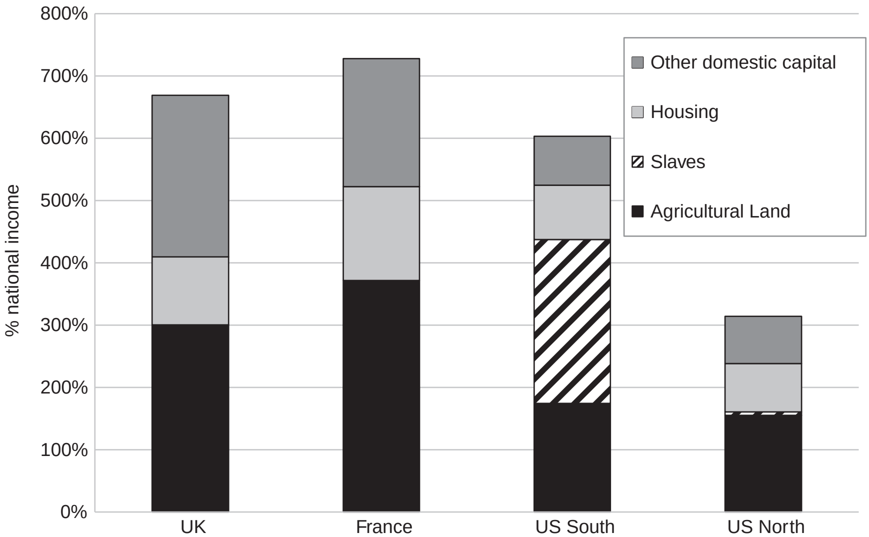


Figure 15.5 National wealth in 1770–1810: Old versus New World.

asset for the owners and a liability for the slaves, so that net national wealth should be unaffected. In the extreme case where a tiny elite owns the rest of the population, the total value of slaves—the total value of "human capital"—could be a lot larger than that of nonhuman capital (since the share of human labor in income is typically larger than

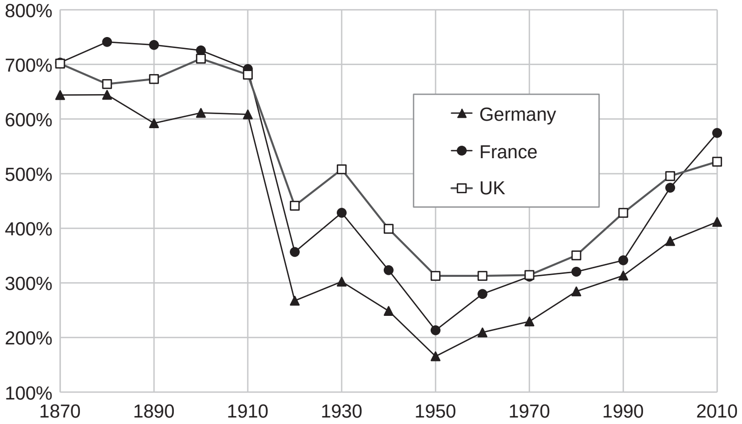
50%). If the rate of return r is equalized across all assets, then the aggregate value of human capital—expressed in proportion to national income—will be equal to $\beta_h = (1-\alpha)/r$, whereas the value of nonhuman capital will be given by $\beta_n = \alpha/r$, where α is the capital share and $1-\alpha$ the labor share implied by the production technology.[12] So for instance with $r=5\%$, $\alpha=30\%$, $1-\alpha=70\%$, the value of the human capital stock will be as large as $\beta_h=(1-\alpha)/r=1400\%$ (14 years of national income), and the value of the nonhuman capital stock will be $\beta_n=\alpha/r=600\%$ (6 years of national income). Outside of slave societies, however, it is unclear whether it makes much sense to compute the market value of human capital and to add it to nonhuman capital.

The computations reported on Figures 15.4 and 15.5 illustrate the ambiguous relationship of the New World with wealth, inequality, and property. To some extent, America is the land of opportunity, the place where wealth accumulated in the past does not matter too much. But it is also the place where a new form of wealth and class structure—arguably more extreme and violent than the class structure prevailing in Europe—flourished, whereby part of the population owned another part.

Available historical series suggest that the sharp U-shaped pattern for the wealth–income ratio in Britain and France is fairly representative of Europe as a whole. For Germany, the wealth–income ratio was approximately the same as for Britain and France in the late nineteenth and early twentieth centuries, then fell to a very low level in the aftermath of the World Wars, and finally has been rising regularly since the 1950s (see Figure 15.6). Although the German wealth–income ratio is still below that of the United Kingdom and France, the speed of the recovery over the past few decades has been similar.[13] On Figure 15.7, we compare the European wealth–income ratio (obtained as a simple average of Britain, France, Germany, and Italy, the latter being available only for the most recent decades) to the U.S. one. The European wealth–income ratio was substantially above that of the United States until World War I, then fell significantly below in the aftermath of World War II, and surpassed it again in the late twentieth and early twenty-first centuries (see Figure 15.7).

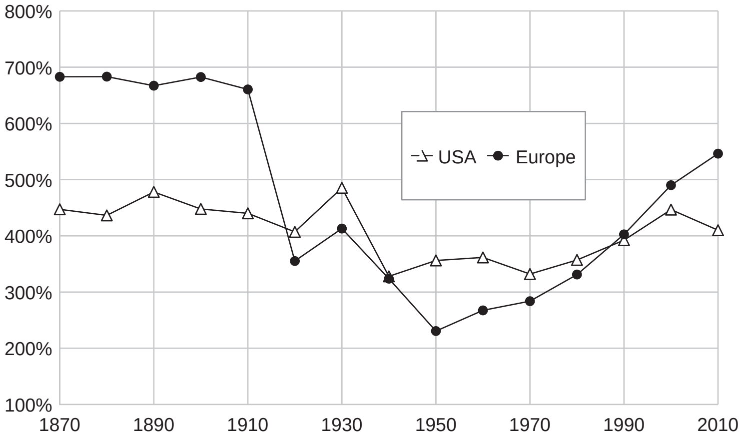
[12] That is, $1-\alpha$ is the marginal product of labor times the labor (slave) stock. The formula $\beta_h=(1-\alpha)/r$ implicitly assumes that the fraction of output that is needed to feed and maintain the slave stock is negligible (otherwise it would just need to be deducted from $1-\alpha$), and that labor productivity is unaffected by the slavery condition (this is a controversial issue).

[13] The factors that can explain the lower German wealth–income ratio are the following. Real estate prices have increased far less in Germany than in Britain or France, which could be due in part to the lasting impact of German reunification and to stronger rent regulations. This could also be temporary. Next, the lower market value of German firms could be due to a stakeholder effect. Finally, the return to the German foreign portfolio, where a large part of German savings were directed, was particularly low in the most recent period. See Piketty and Zucman (2014, Section V.C) and Piketty (2014, Chapter 3).



Private wealth = nonfinancial assets + financial assets–financial liabilities (household & nonprofit sectors).
Data are decennial averages (1910–1913 averages for 1910)

Figure 15.6 Private wealth/national income ratios in Europe, 1870–2010.



Private wealth = nonfinancial assets + financial assets–financial liabilities (household & nonprofit sectors).
Data are decennial averages (1910–1913 averages for Europe)

Figure 15.7 Private wealth/national income ratios 1870–2010: Europe versus United States.

15.2.4 The Return of High Wealth–Income Ratios in Rich Countries

Turning now to the 1970–2010 period, for which we have annual series covering most rich countries, the rise of wealth–income ratios, particularly private wealth–national income ratios, appears to be a general phenomenon. In the top eight developed economies, private wealth is between 2 and 3.5 years of national income around 1970, and between 4 and 7 years of national income around 2010 (see Figure 15.8). Although there are chaotic short-run fluctuations (reflecting the short-run volatility of asset prices),

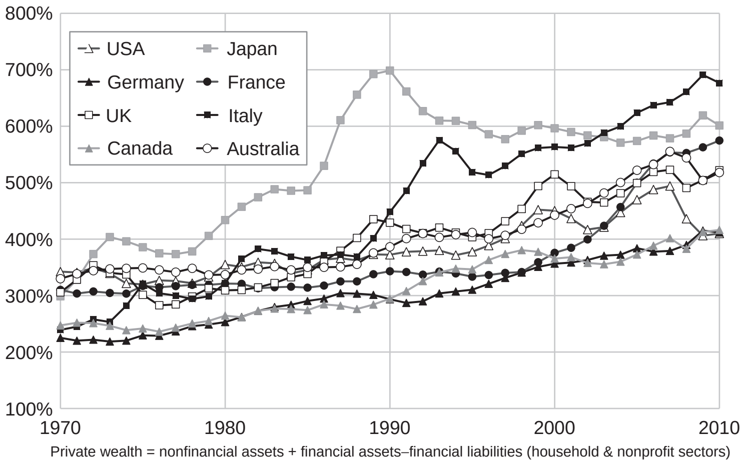


Figure 15.8 Private wealth/national income ratios, 1970–2010.

the long-run trend is clear. Take Japan. The huge asset price bubble of the late 1980s should not obscure the 1970–2010 rise of the wealth–income ratio, fairly comparable in magnitude to what we observe in Europe. (For instance, the Japanese and Italian patterns are relatively close: both countries go from about 2–3 years of national income in private wealth around 1970 to 6–7 years by 2010.)

Although we do not have national wealth estimates for Japan for the late nineteenth and early twentieth centuries, there are reasons to believe that the Japanese wealth–income ratio has also followed a U-shaped evolution in the long run, fairly similar to that observed in Europe over the twentieth century. That is, it seems likely that the wealth–income ratio was relatively high in the early twentieth century, fell to low levels in the aftermath of World War II, and then followed the recovery process that we see in Figure 15.8.[14]

To some extent, the rise of private wealth–national income ratios in rich countries since the 1970s is related to the decline of public wealth (see Figure 15.9). Public wealth has declined virtually everywhere owing both to the rise of public debt and the privatization of public assets. In some countries, such as Italy, public wealth has become strongly negative. The rise in private wealth, however, is quantitatively much larger than the decline in public wealth. As a result, national wealth—the sum of private and public wealth—has increased substantially, from 250–400% of national income in 1970 to 400–650% in 2010 (see Figure 15.10). In Italy, for instance, net government wealth fell by the equivalent of about 1 year of national income, but net private wealth rose by over

[14] The early twentieth century Japanese inheritance tax data reported by Morigushi and Saez (2008) are consistent with this interpretation.

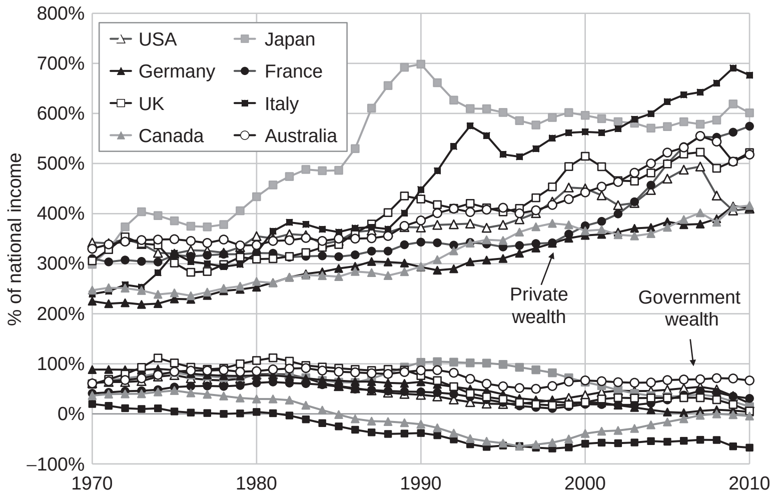


Figure 15.9 Private versus government wealth, 1970–2010.

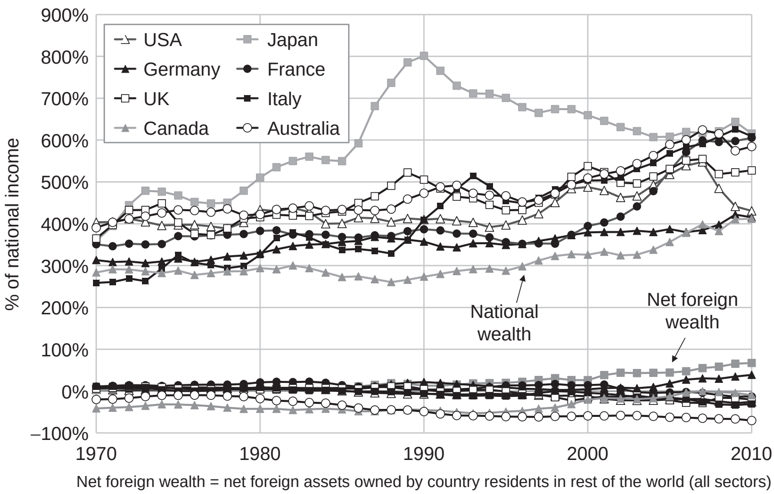


Figure 15.10 National versus foreign wealth, 1970–2010.

4 years of national income, so that national wealth increased by the equivalent of over 3 years of national income.

Figure 15.10 also depicts the evolution of net foreign wealth. Net foreign asset positions are generally small compared to national wealth. In other words, the evolution of national wealth–national income ratios mostly reflects the evolution of domestic capital–output ratios. There are two caveats, however. First, gross cross-border positions have risen a lot in recent decades, which can generate large portfolio valuation effects at the country level. Second, Japan and Germany have accumulated significant net foreign wealth (with net positions around 40% and 70% of national income, respectively, in 2010). Although these are still much smaller than the positions held by France and Britain on the eve of World War I (around 100% and 200% of national income, respectively), they are becoming relatively large (and were rising fast in the case of Germany in the first half of the 2010s, due to the large German trade surpluses).

15.3. THE LONG-RUN EVOLUTION OF WEALTH CONCENTRATION

15.3.1 Concepts, Data Sources, and Methods

We now turn to the evidence on the long-run evolution of wealth concentration. This question can be studied with different data sources (see Davies and Shorrocks, 1999, for a detailed discussion). Ideally, one would want to use annual wealth tax declarations for the entire population. Annual wealth taxes, however, often do to exist, and when they do, the data generally do not cover long periods of time.

The key source used to study the long-run evolution of wealth inequality has traditionally been inheritance and estate tax declarations.[15] By definition, estates and inheritance returns only provide information about wealth at death. The standard way to use inheritance tax data to study wealth concentration was invented over a century ago. Shortly before World War I, a number of British and French economists developed what is known as the mortality multiplier technique, whereby wealth-at-death is weighted by the inverse of the mortality rate of the given age and wealth group in order to generate estimates for the distribution of wealth among the living.[16] This approach was later followed in the United States by Lampman (1962) and Kopczuk and Saez (2004), who use estate tax data covering the 1916–1956 and 1916–2000 periods, respectively, and in the United Kingdom by Atkinson and Harrison (1978), who exploit inheritance tax data covering the 1922–1976 period.

To measure historical trends in the distribution of wealth, one can also use individual income tax returns and capitalize the dividends, interest, rents, and other forms of capital income declared on such returns. The capitalization technique was pioneered by King (1927), Stewart (1939), Atkinson and Harrison (1978), and Greenwood (1983), who used it to estimate the distribution of wealth in the United Kingdom and in the United States for some years in isolation. To obtain reliable results, it is critical to have detailed income data, preferably at the micro level, and to carefully reconcile the tax data with household balance sheets, so as to compute the correct capitalization factors. Drawing

[15] The difference between inheritance and estate taxes is that inheritance taxes are computed at the level of each inheritor, whereas estate taxes are computed at the level of the total estate (total wealth left by the decedent). The raw data coming from these two forms of taxes on wealth transfers are similar.

[16] See Mallet (1908), Séaillès (1910), Strutt (1910), Mallet and Strutt (1915), and Stamp (1919).

on the very detailed U.S. income tax data and Flow of Funds balance sheets, Saez and Zucman (2014) use the capitalization technique to estimate the distribution of U.S. wealth annually since 1913.

For the recent period, one can also use wealth surveys. Surveys, however, are never available on a long-run basis and raise serious difficulties regarding self-reporting biases, especially at the top of the distribution. Tax sources also raise difficulties at the top, especially for the recent period, given the large rise of offshore wealth (Zucman, 2013). Generally speaking, it is certainly more difficult for the recent period to accurately measure the concentration of wealth than the aggregate value of wealth, and one should be aware of this limitation. One needs to be pragmatic and combine the various available data sources (including the global wealth rankings published by magazines such as *Forbes*, which we will refer to in Section 15.5).

The historical series that we analyze in this chapter combines works by many different authors (more details below), who mostly relied on estate and inheritance tax data. They all relate to the inequality of wealth among the living.

We focus on simple concentration indicators, such as the share of aggregate wealth going to the top 10% individuals with the highest net wealth and the share going to the top 1%. In every country and historical period for which we have data, the share of aggregate wealth going to the bottom 50% is extremely small (usually less than 5%). So a decline in the top 10% wealth share can for the most part be interpreted as a rise in the share going to the middle 40%. Note also that wealth concentration is usually almost as large within each age group as for the population taken as a whole.[17]

15.3.2 The European Pattern: France, Britain, and Sweden, 1810–2010

15.3.2.1 France

We start with the case of France, the country for which the longest time series is available. French inheritance tax data is exceptionally good, for one simple reason. As early as 1791, shortly after the abolition of the tax privileges of the aristocracy, the French National Assembly introduced a universal inheritance tax, which has remained in force since then. This inheritance tax was universal because it applied both to bequests and to inter-vivos gifts, at any level of wealth, and for nearly all types of property (both tangible and financial assets). The key characteristic of the tax is that the successors of all decedents with positive wealth, as well as all donees receiving a positive gift, have always been required to file a return, no matter how small the estate was, and no matter whether any tax was ultimately owed.

In other countries, available data are less long run and/or less systematic. In the United Kingdom, one has to wait until 1894 for the unification of inheritance taxation (until this date the rules were different for personal and real estate taxes), and until the early 1920s

[17] See, e.g., Atkinson (1983) and Saez and Zucman (2014).

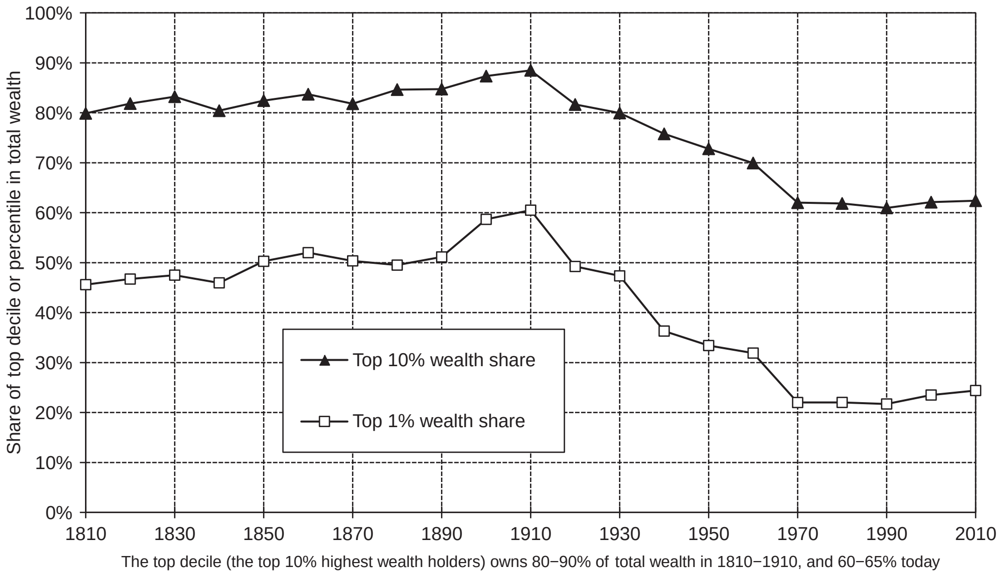


Figure 15.11 Wealth inequality in France, 1810–2010.

for unified statistics to be established by the U.K. tax administration. In the United States, one has to wait until 1916 for the creation of a federal estate tax and the publication of federal statistics on inheritance.

In addition, individual-level inheritance tax declarations have been well preserved in French national archives since the time of the revolution, so that one can use tax registers to collect large representative micro samples. Together with the tabulations by inheritance brackets published by the French tax administration, this allows for a consistent study of wealth inequality over a two-century-long period (see Piketty et al., 2006, 2013).

The main results are summarized on Figures 15.11 and 15.12.[18] First, wealth concentration was very high—and rising—in France during the nineteenth and early twentieth centuries. There was no decline in wealth concentration prior to World War I, quite the contrary: the trend toward rising wealth concentration did accelerate during the 1870–1913 period. The orders of magnitude are quite striking: in 1913, the top 10% wealth share is about 90%, and the top 1% share alone is around 60%. In Paris, which hosts about 5% of the population but as much as 25% of aggregate wealth, wealth is even more concentrated: more than two-thirds of the population has zero or negligible wealth, and 1% of the population owns 70% of the wealth.

[18] The updated series used for Figures 15.11 and 15.12 are based on the historical estimates presented by Piketty et al. (2006) and more recent fiscal data. See Piketty (2014, Chapter 10, Figures 10.1–10.2).

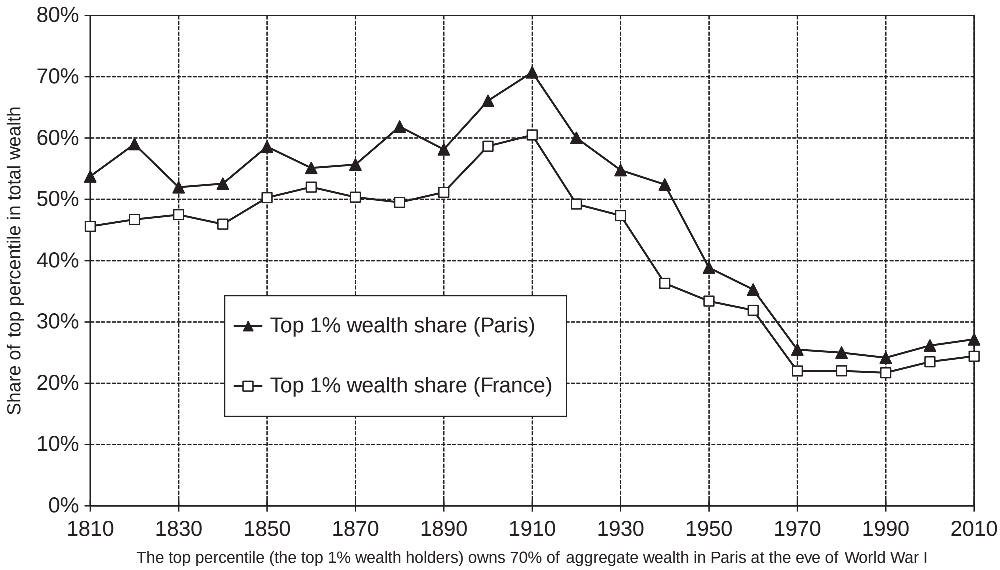


Figure 15.12 Wealth inequality: Paris versus France, 1810–2010.

Looking at Figures 15.11 and 15.12, one naturally wonders whether wealth concentration would have kept increasing without the 1914–1945 shocks. It might have stabilized at a very high level, but it could also have started to decline at some point. In any case, it is clear that the war shocks induced a violent regime change.

The other interesting fact is that wealth concentration has started to increase again in France since the 1970s–1980s—but it is still much lower than on the eve of World War I. According to the most recent data, the top 10% wealth share is slightly above 60%. Given the relatively low quality of today's wealth data, especially regarding top global wealth holders, one should be cautious about this estimate. It could well be that we somewhat underestimate the recent rise and the current level of wealth concentration.[19] In any case, a share of 60% for the top decile is already high, especially compared to the concentration of labor income: the top 10% of labor earners typically receive less than 30% of aggregate labor income.

15.3.2.2 Britain

Although the data sources for other countries are not as systematic and comprehensive as the French sources, existing evidence suggests that the French pattern extends to other European countries. For the United Kingdom, on Figure 15.13, we have combined historical estimates provided by various authors—particularly Atkinson and Harrison (1978)

[19] In contrast, the nineteenth and early twentieth centuries estimates are probably more precise (the tax rates were so low at that time that there was little incentive to hide wealth).

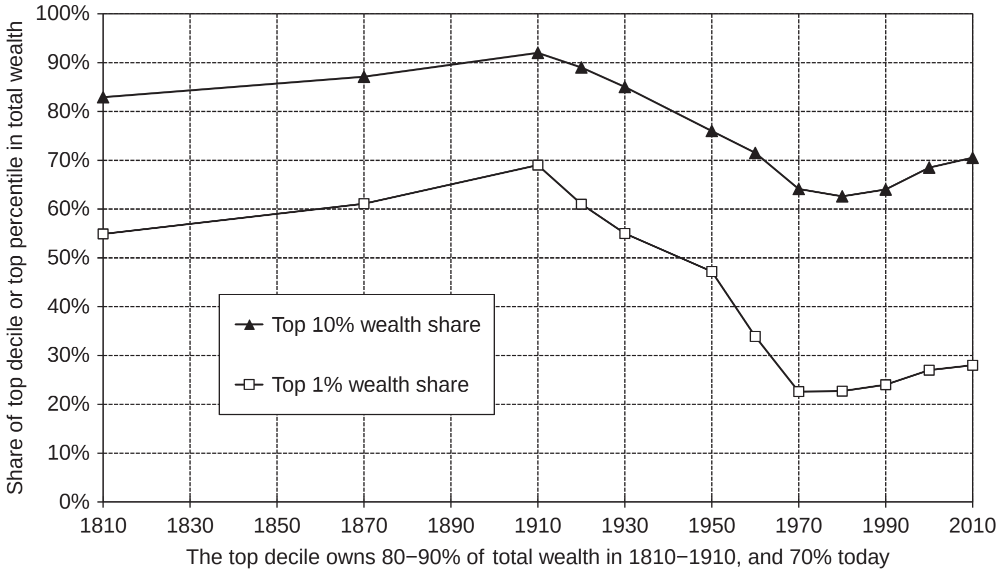


Figure 15.13 Wealth inequality in the United Kingdom, 1810–2010.

and Lindert (1986)—as well as more recent estimates using inheritance tax data. These series are not fully homogenous (in particular, the nineteenth century computations are based on samples of private probate records and are not entirely comparable to the twentieth-century inheritance tax data), but they deliver a consistent picture. Wealth concentration was high and rising during the nineteenth century up until World War I, then fell abruptly following the 1914–1945 shocks, and has been rising again since the 1980s.

According to these estimates, wealth concentration was also somewhat larger in the United Kingdom than in France in the nineteenth and early twentieth centuries. Yet the gap is much smaller than what French contemporary observers claimed. Around 1880–1910, it was very common among French republican elites to describe France as a "country of little property owners" (*un pays de petits propriétaires*), in contrast to aristocratic Britain. Therefore, the argument goes, there was no need to introduce progressive taxation in France (this should be left to Britain). The data show that on the eve of World War I the concentration of wealth was almost as extreme on both sides of the channel: the top 10% owns about 90% of wealth in both countries, and the top 1% owns 70% of wealth in Britain, versus 60% in France. It is true that aristocratic landed estates were more present in the United Kingdom (and to some extent still are today). But given that the share of agricultural land in national wealth dropped to low levels during the nineteenth century (see Figures 15.1 and 15.2), this does not matter much. At the end of the day, whether the country is a republic or a monarchy seems to have little impact on wealth concentration in the long run.

15.3.2.3 Sweden

Although widely regarded as an egalitarian haven today, Sweden was just as unequal as France and Britain in the nineteenth and early twentieth centuries. This is illustrated by Figure 15.13, where we plot some of the estimates constructed by Roine and Waldenstrom (2009) and Waldenstrom (2009).

The concentration of wealth is quite similar across European countries, both for the more ancient and the more recent estimates. Beyond national specificities, a European pattern emerges: the top 10% wealth share went from about 90% around 1900–1910 to about 60–70% in 2000–2010, with a recent rebound. In other words, about 20–30% of national wealth has been redistributed away from the top 10% to the bottom 90%. Since most of this redistribution benefited the middle 40% (the bottom 50% still hardly owns any wealth), this evolution can be described as the rise of a patrimonial middle class (Figure 15.14).

In the case of Sweden, Roine and Waldenstrom (2009) have also computed a corrected top 1% of wealth shares using estimates of offshore wealth held abroad by rich Swedes. They find that under plausible assumptions the top 1% share would shift from about 20% of aggregate wealth to over 30% (i.e., approximately the levels observed in the United Kingdom, and not too far away from the level observed in the United States). This illustrates the limitations of our ability to measure recent trends and levels, given the rising importance of tax havens.

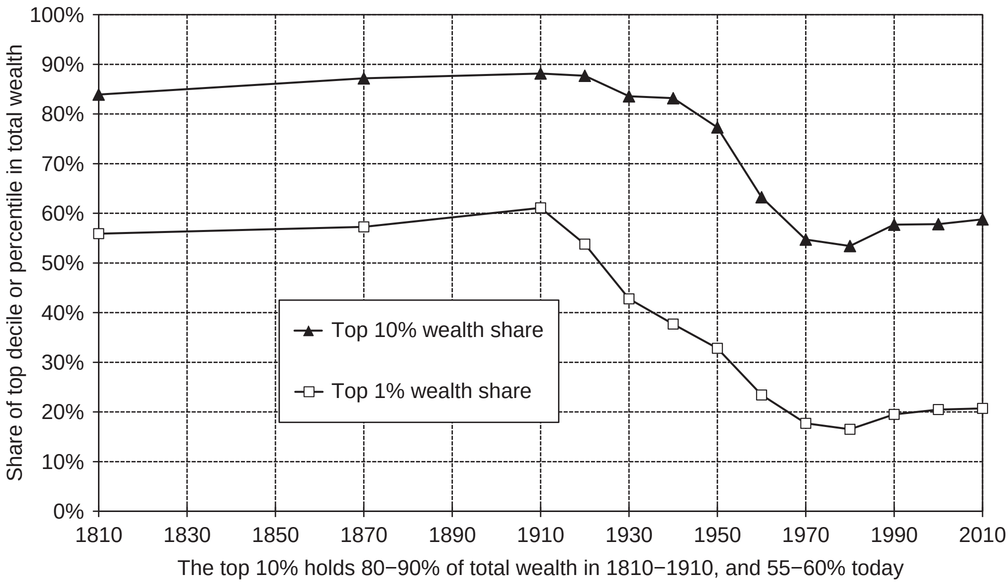


Figure 15.14 Wealth inequality in Sweden, 1810–2010.

15.3.3 The Great Inequality Reversal: Europe Versus the United States, 1810–2010

Comparing wealth concentration in Europe and the United States, the main finding is a fairly spectacular reversal. In the nineteenth century, the United States was to some extent the land of equality (at least for white men): the concentration of wealth was much less extreme than in Europe (except in the South). Over the course of the twentieth century, this ordering was reversed: wealth concentration has become significantly higher in the United States. This is illustrated by Figure 15.15, where we combine the estimates due to Lindert (2000) for the nineteenth century with those of Saez and Zucman (2014) for the twentieth and twenty-first centuries to form long-run U.S. series, and by Figure 15.16, where we compare the United States to Europe (defined as the arithmetic average of France, Britain, and Sweden).

The reversal comes from the fact that Europe has become significantly less unequal over the course of the twentieth century, whereas the United States has not. The United States has almost returned to its early twentieth-century wealth concentration level: at its peak in the late 1920s, the 10% wealth share was about 80%, in 2012 it is about 75%; similarly the top 1% share peaked at about 45% and is back to around 40% today. Note, however, that the United States never reached the extreme level of wealth concentration of nineteenth- and early twentieth-century Europe (with a top decile of 90% or more). The United States has always had a patrimonial middle class, although one of varying importance. The share of wealth held by the middle class appears to have been shrinking since the 1980s.

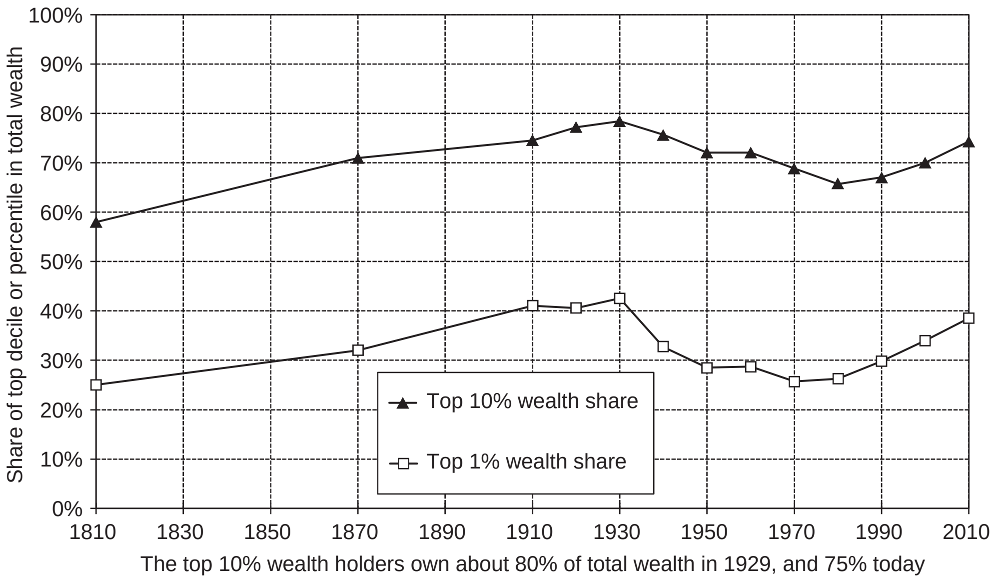


Figure 15.15 Wealth inequality in the United States, 1810–2010.

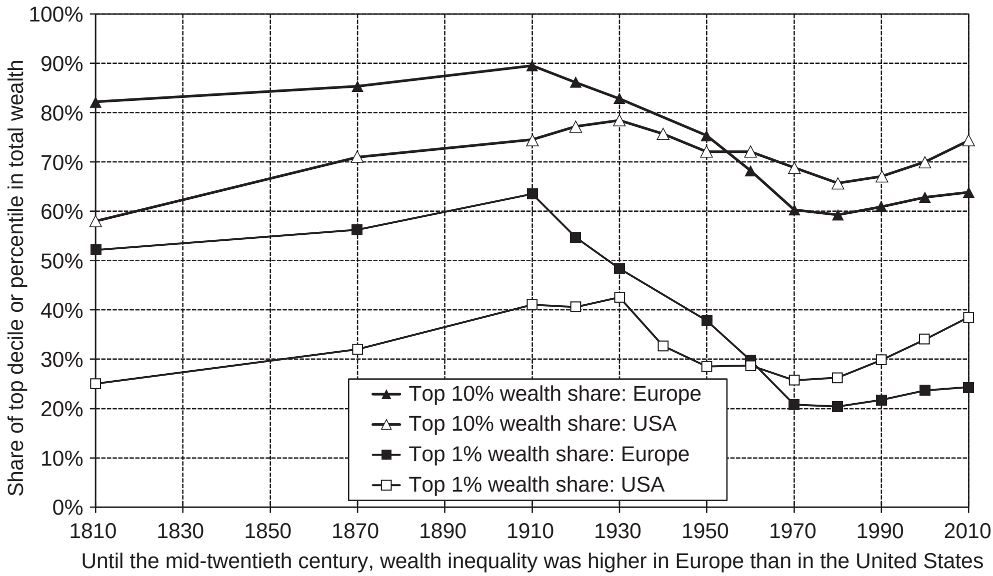


Figure 15.16 Wealth inequality: Europe and the United States, 1810–2010.

U.S. economists of the early twentieth century were very concerned about the possibility that their country becomes as unequal as Old Europe. Irving Fisher, then president of the American Economic Association, gave his presidential address in 1919 on this topic. He argued that the concentration of income and wealth was becoming as dangerously excessive in America as it had been for a long time in Europe. He called for steep tax progressivity to counteract this tendency. Fisher was particularly concerned about the fact that as much as half of U.S. wealth was owned by just 2% of U.S. population, a situation that he viewed as "undemocratic" (see Fisher, 1920). One can indeed interpret the spectacular rise of tax progressivity that occurred in the United States during the first half of the twentieth century as an attempt to preserve the egalitarian, democratic American ethos (celebrated a century before by Tocqueville and others). Attitudes toward inequality are dramatically different today. Many U.S. observers now view Europe as excessively egalitarian (and many European observers view the United States as excessively nonegalitarian).

15.4. THE LONG-RUN EVOLUTION OF THE SHARE OF INHERITED WEALTH

15.4.1 Concepts, Data Sources, and Methods

We now turn to our third ratio of interest, the share of inherited wealth in aggregate wealth. We should make clear at the outset that this is an area where available evidence is scarce and incomplete. Measuring the share of inherited wealth requires a lot more data than the measurement of aggregate wealth–income ratios or even wealth concentration. It is also an area where it is important to be particularly careful about concepts and definitions. Purely definitional conflicts have caused substantial confusion in the past. Therefore it is critical to start from there.

15.4.1.1 Basic Notions and Definitions

The most natural way to define the share of inherited wealth in aggregate wealth is to cumulate past inheritance flows. That is, assume that we observe the aggregate wealth stock W_t at time t in a given country, and that we would like to define and estimate the aggregate inherited wealth stock $W_{Bt} \leq W_t$ (and conversely aggregate self-made wealth, which we simply define as $W_{St} = W_t - W_{Bt}$). Assume that we observe the annual flow of inheritance B_s that occurred in any year $s \leq t$. At first sight, it might seem natural to define the stock of inherited wealth W_{Bt} as the sum of past inheritance flows:

$$W_{Bt} = \int_{s \leq t} B_s \cdot \mathrm{d}s$$

However, there are several practical and conceptual difficulties with this ambiguous definition, which need to be addressed before the formula can be applied to actual data. First, it is critical to include in this sum not only past bequest flows B_s (wealth transmissions at death) but also inter vivos gift flows V_s (wealth transmissions inter vivos). That is, one should define W_{Bt} as $W_{Bt} = \int_{s \leq t} B_s^{*} \cdot \mathrm{d}s$, with $B_s^{*} = B_s + V_s$.

Alternatively, if one cannot observe directly the gift flow V_s, one should replace the observed bequest flow B_s by some gross level $B_s^{*} = (1 + v_s) \cdot B_s$, where $v_s = V_s / B_s$ is an estimate of the gift/bequest flow ratio. In countries where adequate data is available, the gift–bequest ratio is at least 10–20%, and is often higher than 50%, especially in the recent period.[20] It is thus critical to include gifts in one way or another. In countries where fiscal data on gifts are insufficient, one should at least try to estimate $1 + v_s$ using surveys (which often suffers from severe downward biases) and harder administrative evidence from other countries.

Next, to properly apply this definition, one should only take into account the fraction of the aggregate inheritance flow $B_{st} \leq B_s$ that was received at time s by individuals who are still alive at time t. The problem is that doing so properly requires very detailed individual-level information. At any time t, there are always individuals who received inheritance a very long time ago (say, 60 years ago) but who are still alive (because they inherited at a very young age and/or are enjoying a very long life). Conversely, a fraction

[20] See below. Usually one only includes formal, monetary capital gifts, and one ignores informal presents and in-kind gifts. In particular in-kind gifts made to minors living with their parents (i.e., the fact that minor children are usually catered by their parents) are generally left aside.

of the inheritance flow received a short time ago (say, 10 years ago) should not be counted (because the relevant inheritors are already dead, e.g., they inherited at an old age or died young). In practice, however, such unusual events tend to balance each other, so that a standard simplifying assumption is to cumulate the full inheritance flows observed the previous H years, where H is the average generation length, that is, the average age at which parents have children (typically $H = 30$ years). Therefore we obtain the following simplified definition:

$$W_{Bt} = \int_{t-30 \leq s \leq t} (1 + v_s) \cdot B_s \cdot \mathrm{d}s$$

15.4.1.2 The Kotlikoff–Summers–Modigliani Controversy

Assume now that these two difficulties can be addressed (i.e., that we can properly estimate the factor $1 + v_s$ and the average generation length H). There are more substantial difficulties ahead. First, to properly compute W_{Bt}, one needs to be able to observe inheritance flows B_s^* over a relatively long time period (typically, the previous 30 years). In the famous Kotlikoff–Summers–Modigliani (KSM) controversy, both Kotlikoff and Summers (1981) and Modigliani (1986, 1988) used estimates of the U.S. inheritance flow for only 1 year (and a relatively ancient year: 1962), see also Kotlikoff (1988). They simply assumed that this estimate could be used for other years. Namely, they assumed that the inheritance flow–national income ratio (which we note $b_{ys} = B_s^*/Y_s$) is stable over time. One problem with this assumption is that it might not be verified. As we shall see below, extensive historical data on inheritances recently collected in France show that the b_{ys} ratio has changed tremendously over the past two centuries, from about 20–25% of national income in the nineteenth and early twentieth centuries, down to less than 5% at mid-twentieth century, back to about 15% in the early twenty-first century (Piketty, 2011). So one cannot simply use one year of data and assume that we are in a steady state: One needs a long-run time series on the inheritance flow in order to estimate the aggregate stock of inherited wealth.

Next, one needs to decide the extent to which past inheritance flows need to be upgraded or capitalized. This is the main source of disagreement and confusion in the KSM controversy.

Modigliani (1986, 1988) chooses zero capitalization. That is, he simply defines the stock of inherited wealth W_{Bt}^M as the raw sum of past inheritance flows with no adjustment whatsoever (except for the GDP price index):

$$W_{Bt}^M = \int_{t-30 \leq s \leq t} B_s^* \cdot \mathrm{d}s$$

Assume a fixed inheritance flow–national income ratio $b_y = B_s^*/Y_s$, growth rate g (so that $Y_t = Y_s \cdot e^{g(t-s)}$), generation length H, and aggregate private wealth–national income ratio

$\beta = W_t/Y_t$. Then, according to the Modigliani definition, the steady-state formulas for the stock of inherited wealth relative to national income W_{Bt}^M/Y_t and for the share of inherited wealth $\varphi_t^M = W_{Bt}^M/W_t$ are given by

$$W_{Bt}^M/Y_t = \frac{1}{Y_t}\int_{t-30\leq s\leq t} B_s^*\cdot \mathrm{d}s = \frac{1-\mathrm{e}^{-gH}}{g}\cdot b_y$$

$$\varphi_t^M = W_{Bt}^M/W_t = \frac{1-\mathrm{e}^{-gH}}{g}\cdot\frac{b_y}{\beta}$$

In contrast, Kotlikoff and Summers (1981, 1988) choose to capitalize past inheritance flows by using the economy's average rate of return to wealth (assuming it is constant and equal to r). Following the Kotlikoff–Summers definition, the steady-state formulas for the stock of inherited wealth relative to national income W_{Bt}^{KS}/Y_t and for the share of inherited wealth $\varphi_t^{KS} = W_{Bt}^{KS}/W_t$ are given by

$$W_{Bt}^{KS}/Y_t = \frac{1}{Y_t}\int_{t-30\leq s\leq t} \mathrm{e}^{r(t-s)}\cdot B_s^*\cdot \mathrm{d}s = \frac{\mathrm{e}^{(r-g)H}-1}{r-g}\cdot b_y$$

$$\varphi_t^{KS} = W_{Bt}^{KS}/W_t = \frac{\mathrm{e}^{(r-g)H}-1}{r-g}\cdot\frac{b_y}{\beta}$$

In the special case where growth rates and rates of return are negligible (i.e., infinitely close to zero), then both definitions coincide. That is, if $g=0$ and $r-g=0$, then $(1-\mathrm{e}^{-gH})/g = (\mathrm{e}^{(r-g)H}-1)/(r-g) = H$, so that $W_{Bt}^M/Y_t = W_{Bt}^{KS}/Y_t = Hb_y$ and $\varphi_t^M = \varphi_t^{KS} = Hb_y/\beta$.

Thus, in case growth and capitalization effects can be neglected, one simply needs to multiply the annual inheritance flow by generation length. If the annual inheritance flow is equal to $b_y = 10\%$ of national income, and generation length is equal to $H = 30$ years, then the stock of inherited wealth is equal to $W_{Bt}^M = W_{Bt}^{KS} = 300\%$ of national income according to both definitions. In case aggregate wealth amounts to $\beta = 400\%$ of national income, then the inheritance share is equal to $\varphi_t^M = \varphi_t^{KS} = 75\%$ of aggregate wealth.

However, in the general case where g and $r-g$ are significantly different from zero, the two definitions can lead to widely different conclusions. For instance, with $g = 2\%$, $r = 4\%$, and $H = 30$, we have the following capitalization factors: $(1-\mathrm{e}^{-gH})/(g\cdot H) = 0.75$ and $(\mathrm{e}^{(r-g)H}-1)/((r-g)\cdot H) = 1.37$. In this example, for a given inheritance flow $b_y = 10\%$ and aggregate wealth–income ratio $\beta = 400\%$, we obtain $\varphi_t^M = 56\%$ and $\varphi_t^{KS} = 103\%$. About half of wealth comes from inheritance according to the Modigliani definition, and all of it according to the Kotlikoff–Summers definition.

This is the main reason why Modigliani and Kotlikoff–Summers disagree so much about the inheritance share. They both use the same (relatively fragile) estimate for the United States b_y in 1962. But Modigliani does not capitalize past inheritance flows and concludes that the inheritance share is as low as 20–30%. Kotlikoff–Summers do capitalize the same flows and conclude that the inheritance share is as large as 80–90% (or even larger than 100%). Both sides also disagree somewhat about the measurement of b_y, but the main source of disagreement comes from this capitalization effect.[21]

15.4.1.3 The Limitations of KSM Definitions

Which of the two definitions is most justified? In our view, both are problematic. It is wholly inappropriate not to capitalize at all past inheritance flows. But full capitalization is also inadequate.

The key problem with the KSM representative-agent approach is that it fails to recognize that the wealth accumulation process always involves two different kinds of people and wealth trajectories. In every economy, there are inheritors (people who typically consume part of the return to their inherited wealth) and there are savers (people who do not inherit much but do accumulate wealth through labor income savings). This is an important feature of the real world that must be taken into account for a proper understanding of the aggregate wealth accumulation process.

The Modigliani definition is particularly problematic as it simply fails to recognize that inherited wealth produces flow returns. This mechanically leads to artificially low numbers for the inheritance share φ_t^M (as low as 20–40%), and to artificially high numbers for the life-cycle share in wealth accumulation, which Modigliani defines as $1-\varphi_t^M$ (up to 60–80%). As Blinder (1988) argues, "a Rockefeller with zero lifetime labor income and consuming only part of his inherited wealth income would appear to be a life-cycle saver in Modigliani's definition, which seems weird to me." One can easily construct illustrative examples of economies where all wealth comes from inheritance (with dynasties of the sort described by Blinder), but where Modigliani would still find an inheritance share well below 50%, simply because of his definition. This makes little sense.[22]

The Kotlikoff–Summers definition is conceptually more satisfactory than Modigliani's. But it suffers from the opposite drawback in the sense that it mechanically leads to artificially high numbers for the inheritance share φ_t^{KS}. In particular, φ_t^{KS} can easily be larger than 100%, even though there are life-cycle savers and self-made wealth accumulators in the economy, and a significant fraction of aggregate wealth accumulation comes from them. This will arise whenever the cumulated return to inherited wealth

[21] In effect, Modigliani favors a b_y ratio around 5–6%, whereas Kotlikoff–Summers find it more realistic to use a b_y ratio around 7–8%. Given the data sources they use, it is likely that both sides tend to somewhat underestimate the true ratio. See the discussion below for the case of France and other European countries.

[22] It is worth stressing that the return to inherited wealth (and the possibility to save and accumulate more wealth out of the return to inherited wealth) is a highly relevant economic issue not only for high-wealth dynasties of the sort referred to by Blinder, but also for middle-wealth dynasties. For instance, it is easier to save if one has inherited a house and has no rent to pay. An inheritor saving less than the rental value of his inherited home would be described as a life-cycle saver according to Modigliani's definition, which again seems odd.

consumed by inheritors exceeds the savers' wealth accumulation from their labor savings. In the real world, this condition seems to hold not only in prototype rentier societies such as Paris 1872–1937 (see Piketty et al., 2013), but also in countries and time periods when aggregate inheritance flow is relatively low. For instance, aggregate French series show that the capitalized bequest share φ_t^{KS} has been larger than 100% throughout the twentieth century, including in the 1950s–1970s, a period where a very significant amount of new self-made wealth was accumulated (Piketty, 2011).

In sum, the Modigliani definition leads to estimates of the inheritance share that are artificially close to 0%, whereas the Kotlikoff–Summers leads to inheritance shares that tend to be structurally above 100%. Neither of them offers an adequate way to look at the data.

15.4.1.4 The PPVR Definition

In an ideal world with perfect data, the conceptually consistent way to define the share of inherited wealth in aggregate wealth is the following. It has first been formalized and applied to Parisian wealth data by Piketty et al. (2013), so we refer to it as the PPVR definition.

The basic idea is to split the population into two groups. First, there are "inheritors" (or "rentiers"), whose assets are worth less than the capitalized value of the wealth they inherited (over time they consume more than their labor income). The second group is composed of "savers" (or "self-made individuals"), whose assets are worth more than the capitalized value of the wealth they inherited (they consume less than their labor income). Aggregate inherited wealth can then be defined as the sum of inheritors' wealth plus the inherited fraction of savers' wealth, and self-made wealth as the noninherited fraction of savers' wealth. By construction, inherited and self-made wealth are less than 100% and sum to aggregate wealth, which is certainly a desirable property. Although the definition is fairly straightforward, it differs considerably from the standard KSM definitions based on representative agent models. The PPVR definition is conceptually more consistent and provides a more meaningful way to look at the data and to analyze the structure of wealth accumulation processes. In effect, it amounts to defining inherited wealth at the individual level as the minimum between current wealth and capitalized inheritance.

More precisely, consider an economy with population N_t at time t. Take a given individual i with wealth w_{ti} at time t. Assume he or she received bequest b_{ti}^0 at time $t_i < t$. Note $b_{ti}^* = b_{ti}^0 \cdot e^{r(t-t_i)}$ the capitalized value of b_{ti}^0 at time t (where $e^{r(t-ti)}$ is the cumulated rate of return between time t_i and time t). Individual i is said to be an "inheritor" (or a "rentier") if $w_{ti} < b_{ti}^*$ and a "saver" (or a "self-made individual") if $w_{ti} \geq b_{ti}^*$. We define the set of inheritors as $N_t^r = \{i \,\mathrm{s.t.}\, w_{ti} < b_{ti}^*\}$ and the set of savers as $N_t^s = \{i \,\mathrm{s.t.}\, w_{ti} \geq b_{ti}^*\}$.

We note $\rho_t = N_t^r/N_t$ and $1-\rho_t = N_t^s/N_t$ as the corresponding population shares of inheritors and savers; $w_t^r = E(w_{ti}|w_{ti} < b_{ti}^*)$ and $w_t^s = E(w_{ti}|w_{ti} \geq b_{ti}^*)$ as the average wealth levels of both groups; $br_t^* = E(b_{ti}^*|w_{ti} < b_{ti}^*)$ and $bs_t^* = E(b_{ti}^*|w_{ti} \geq b_{ti}^*)$ as the levels of their average capitalized bequest; and $\pi_t = \rho_t \cdot w_t^r / w_t$ and $1-\pi_t = (1-\rho_t)\cdot w_t^s / w_t$ as the share of inheritors and savers in aggregate wealth.

We define the total share φ_t of inherited wealth in aggregate wealth as the sum of inheritors' wealth plus the inherited fraction of savers' wealth, and the share $1-\varphi_t$ of self-made wealth as the noninherited fraction of savers' wealth:

$$\varphi_t = \left[\rho_t \cdot w_t^r + (1-\rho_t)\cdot b_t^{s*}\right]/w_t = \pi_t + (1-\rho_t)\cdot b_t^{s*}/w_t$$

$$1-\varphi_t = (1-\rho_t)\cdot\left(w_t^s - b_t^{s*}\right)/w_t = 1-\pi_t-(1-\rho_t)\cdot b_t^{s*}/w_t$$

The downside of this definition is that it is more demanding in terms of data availability. Although Modigliani and Kotlikoff–Summers could compute inheritance shares in aggregate wealth by using aggregate data only, the PPVR definition requires micro data. Namely, we need data on the joint distribution $G_t(w_{ti}, b_{ti}^*)$ of current wealth w_{ti} and capitalized inherited wealth b_{ti}^* in order to compute ρ_t, π_t, and φ_t. This does require high-quality, individual-level data on wealth and inheritance over two generations, which is often difficult to obtain. It is worth stressing, however, that we do not need to know anything about the individual labor income or consumption paths (y_{Lsi}, c_{si}, $s<t$) followed by individual i up to the time of observation.[23]

For plausible joint distributions $G_t(w_{ti}, b_{ti}^*)$, the PPVR inheritance share φ_t will typically fall somewhere in the interval $[\varphi_t^M, \varphi_t^{KS}]$. There is, however, no theoretical reason why it should be so in general. Imagine, for instance, an economy where inheritors consume their bequests the very day they receive them, and never save afterward, so that wealth accumulation entirely comes from the savers, who never received any bequest (or negligible amounts) and who patiently accumulate savings from their labor income. Then with our definition $\varphi_t = 0\%$: in this economy, 100% of wealth accumulation comes from savings, and nothing at all comes from inheritance.

However, with the Modigliani and Kotlikoff–Summers definitions, the inheritance shares φ_t^M and φ_t^{KS} could be arbitrarily large.

[23] Of course, more data are better. If we also have (or estimate) labor income or consumption paths, then one can compute lifetime individual savings rate s_{Bti}, that is, the share of lifetime resources that was not consumed up to time t: $s_{Bti} = w_{ti}/(b_{ti}^* + y_{Lti}^*) = 1 - c_{ti}^*/(b_{ti}^* + y_{Lti}^*)$ with $y_{Lti}^* = \int_{s<t} y_{Lsi} e^{r(t-s)} ds$ = capitalized value at time t of past labor income flows, and $c_{ti}^* = \int_{s<t} c_{si} e^{r(t-s)} ds$ = capitalized value at time t of past consumption flows. By definition, inheritors are individuals who consumed more than their labor income (i.e., $w_{ti} < b_{ti}^* \leftrightarrow c_{ti}^* > y_{Lti}^*$), while savers are individuals who consumed less than their labor income (i.e., $w_{ti} \geq b_{ti}^* \leftrightarrow c_{ti}^* \leq y_{Lti}^*$). But the point is that we only need to observe an individual's wealth (w_{ti}) and capitalized inheritance (b_{ti}^*) to determine whether he or she is an inheritor or a saver, and in order to compute the share of inherited wealth.

15.4.1.5 A Simplified Definition: Inheritance Flows versus Saving Flows

When available micro data is not sufficient to apply the PPVR definition, one can also use a simplified, approximate definition based on the comparison between inheritance flows and saving flows.

Assume that all we have is macro data on inheritance flows $b_{yt} = B_t / Y_t$ and savings flows $s_t = S_t / Y_t$. Suppose for simplicity that both flows are constant over time: $b_{yt} = b_y$ and $s_t = s$. We want to estimate the share $\varphi = W_B / W$ of inherited wealth in aggregate wealth. The difficulty is that we typically do not know which part of the aggregate saving rate s comes from the return to inherited wealth, and which part comes from labor income (or from the return to past savings). Ideally, one would like to distinguish between the savings of inheritors and self-made individuals (defined along the lines explained above), but this requires micro data over two generations. In the absence of such data, a natural starting point would be to assume that the propensity to save is on average the same whatever the income sources. That is, a fraction $\varphi \cdot \alpha$ of the saving rate s should be attributed to the return to inherited wealth, and a fraction $1 - \alpha + (1 - \varphi) \cdot \alpha$ should be attributed to labor income (and to the return to past savings), where $\alpha = Y_K / Y$ is the capital share in national income and $1 - \alpha = Y_L / Y$ is the labor share. Assuming again that we are in steady state, we obtain the following simplified formula for the share of inherited wealth in aggregate wealth:

$$\varphi = \frac{b_y + \varphi \cdot \alpha \cdot s}{b_y + s},$$

$$\text{i.e.,} \varphi = \frac{b_y}{b_y + (1 - \alpha) \cdot s}.$$

Intuitively, this formula simply compares the size of the inheritance and saving flows. Because all wealth must originate from one of the two flows, it is the most natural way to estimate the share of inherited wealth in total wealth.[24]

There are a number of caveats with this simplified formula. First, real-world economies are generally out of steady state, so it is important to compute average values of b_y, s, and α over relatively long periods of time (typically over the past H years, with $H = 30$ years). If one has time-series estimates of the inheritance flow b_{ys}, capital share α_s, and saving rate s_s, then one can use the following full formula, which capitalizes past inheritance and savings flows at rate $r - g$:

[24] Similar formulas based on the comparison of inheritance and saving flows have been used by DeLong (2003) and Davies et al. (2012, pp. 123–124). One important difference is that these authors do not take into account the fact that the savings flow partly comes from the return to inherited wealth. We return to this point in Section **15.5.4**.

$$\varphi = \frac{\int_{t-H\le s\le t} e^{(r-g)(t-s)}\cdot b_{ys}\cdot ds}{\int_{t-H\le s\le t} e^{(r-g)(t-s)}\cdot\left(b_{ys}+(1-\alpha_s)\cdot s_s\right)\cdot ds}.$$

With constant flows, the full formula boils down to $\varphi = \frac{b_y}{b_y+(1-\alpha)\cdot s}$.

Second, one should bear in mind that the simplified formula $\varphi = b_y/(b_y+(1-\alpha)\cdot s)$ is an approximate formula. In general, as we show below, it tends to underestimate the true share of inheritance, as computed from micro data using the PPVR definition. The reason is that individuals who have only labor income tend to save less (in proportion to their total income) than those who have large inherited wealth and capital income, which. in turn, seems to be related to the fact that wealth (and particularly inherited wealth) is more concentrated than labor income.

On the positive side, simplified estimates of φ seem to follow micro-based estimates relatively closely (much more closely than KSM estimates, which are either far too small or far too large), and they are much less demanding in terms of data. One only needs to estimate macro flows. Another key advantage of the simplified definition over KSM definitions is that it does not depend upon the sensitive choice of the rate of return or the rate of capital gains or losses. Whatever these rates might be, they should apply equally to inherited and self-made wealth (at least as a first approximation), so one can simply compare inheritance and saving flows.

15.4.2 The Long-Run Evolution of Inheritance in France 1820–2010

15.4.2.1 The Inheritance Flow–National Income Ratio b_{yt}

What do we empirically know about the historical evolution of inheritance? We start by presenting the evidence on the dynamics of the inheritance to national income ratio b_{yt} in France, a country for which, as we have seen in Section 15.3, historical data sources are exceptionally good (Piketty, 2011). The main conclusion is that b_{yt} has followed a spectacular U-shaped pattern over the twentieth century. The inheritance flow was relatively stable, around 20–25% of national income throughout the 1820–1910 period (with a slight upward trend), before being divided by a factor of about 5–6 between 1910 and the 1950s, and then multiplied by a factor of about 3–4 between the 1950s and the 2000s (see Figure 15.17).

These are enormous historical variations, but they appear to be well-founded empirically. In particular, the patterns for b_{yt} are similar with two independent measures of the inheritance flow. The first, what we call the fiscal flow, uses bequest and gift tax data and makes allowances for tax-exempt assets such as life insurance. The second measure, what we call the economic flow, combines estimates of private wealth W_t, mortality tables, and observed age–wealth profile, using the following accounting equation:

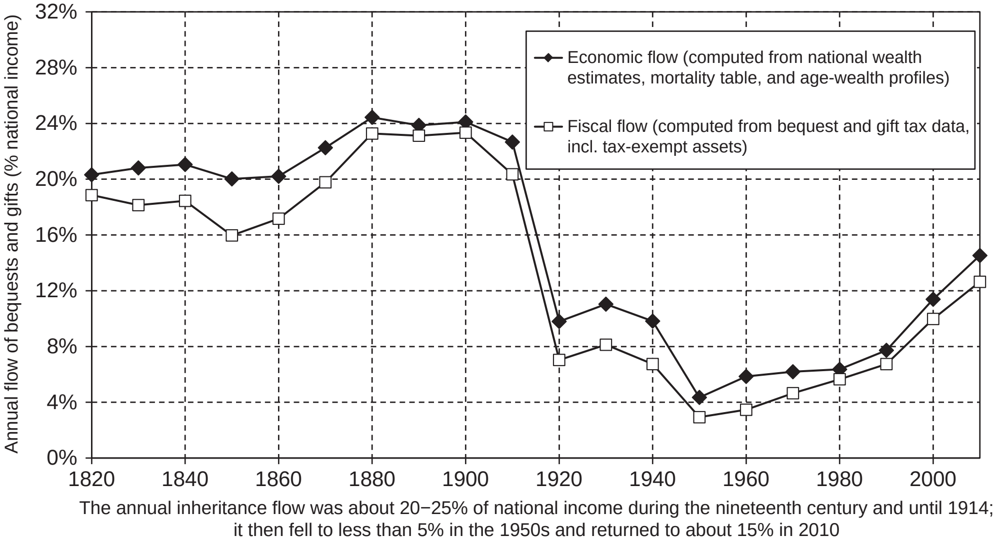


Figure 15.17 The annual inheritance flow as a fraction of national income, France 1820–2010.

$$B_t^* = (1 + v_t) \cdot \mu_t \cdot m_t \cdot W_t,$$

where m_t = mortality rate (number of adult decedents divided by total adult population), μ_t = ratio between average adult wealth at death and average adult wealth for the entire population, and $v_t = V_t / B_t$ = estimate of the gift/bequest flow ratio.

The gap between the fiscal and economic flows can be interpreted as capturing tax evasion and other measurement errors. It is approximately constant over time and relatively small, so that the two series deliver consistent long-run patterns (see Figure 15.17).

The economic flow series allow—by construction—for a straightforward decomposition of the various effects at play in the evolution of b_{yt}. In the above equation, dividing both terms by Y_t we get

$$b_{yt} = B_t^* / Y_t = (1 + v_t) \cdot \mu_t \cdot m_t \cdot \beta_t.$$

Similarly, dividing by W_t we can define the rate of wealth transmission b_{wt} as

$$\begin{aligned} b_{wt} &= B_t^* / W_t = (1 + v_t) \cdot \mu_t \cdot m_t = \mu_t^* \cdot m_t \\ \text{with} \, \mu_t^* &= (1 + v_t) \cdot \mu_t = \text{gift-corrected ratio} \end{aligned}$$

If $\mu_t = 1$ (i.e., decedents have the same average wealth as the living) and $v_t = 0$ (no gift), then the rate of wealth transmission is simply equal to the mortality rate: $b_{wt} = m_t$ (and $b_{yt} = m_t \cdot \beta_t$). If $\mu_t = 0$ (i.e., decedents die with zero wealth, such as in Modigliani's pure life-cycle theory of wealth accumulation) and $v_t = 0$ (no gift), then there is no inheritance at all: $b_{wt} = b_{yt} = 0$.

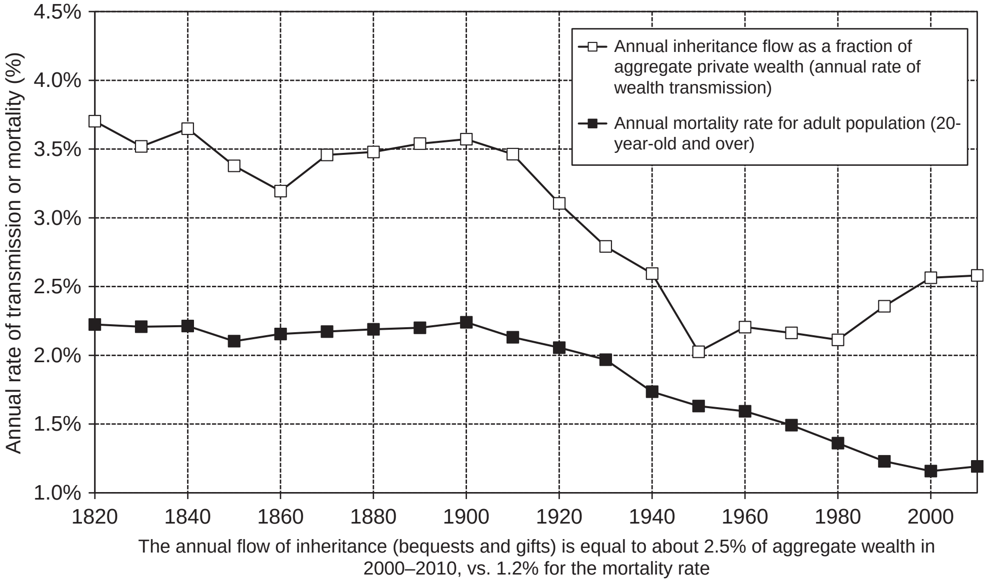


Figure 15.18 Inheritance flow versus mortality rate, France 1820–2010.

Using these accounting equations, we can see that the U-shaped pattern followed by the French inheritance-income ratio b_{yt} is the product of two U-shaped evolutions. First, it partly comes from the U-shaped evolution of the private wealth–income ratio β_t. The U-shaped evolution of b_{yt}, however, is almost twice as marked at that of β_t. The wealth–income ratio was divided by a factor of about 2–3 between 1910 and 1950 (from 600–700% to 200–300%, see Figure 15.2), whereas the inheritance flow was divided by a factor around 5–6 (from 20–25% to about 4%, see Figure 15.17). The explanation is that the rate of wealth transmission $b_{wt} = \mu_t^* \cdot m_t$ has also been following a U-shaped pattern: it was almost divided by 2 between 1910 and 1950 (from over 3.5% to just 2%), and it has been rising again to about 2.5% in 2010 (see Figure 15.18).

The U-shaped pattern followed by b_{wt}, in turn, entirely comes from μ_t^*. The relative wealth of decedents was at its lowest historical level in the aftermath of World War II (which, as we shall see below, is largely due to the fact that it was too late for older cohorts to recover from the shocks and reaccumulate wealth after the war). Given that aggregate wealth was also at its lowest historical level, the combination of these two factors explains the exceptionally low level of the inheritance flow in the 1950s–1960s. By contrast, the mortality rate m_t has been constantly diminishing: this long-run downward trend is the mechanical consequence of the rise in life expectancy (for a given cohort size).[25]

[25] The mortality rate, however, is about to rise somewhat in coming decades in France owing to the aging of the baby boomers (see Piketty, 2011). This effect will be even stronger in countries where cohort size has declined in recent decades (such as Germany or Japan) and will tend to push inheritance flows toward even higher levels.

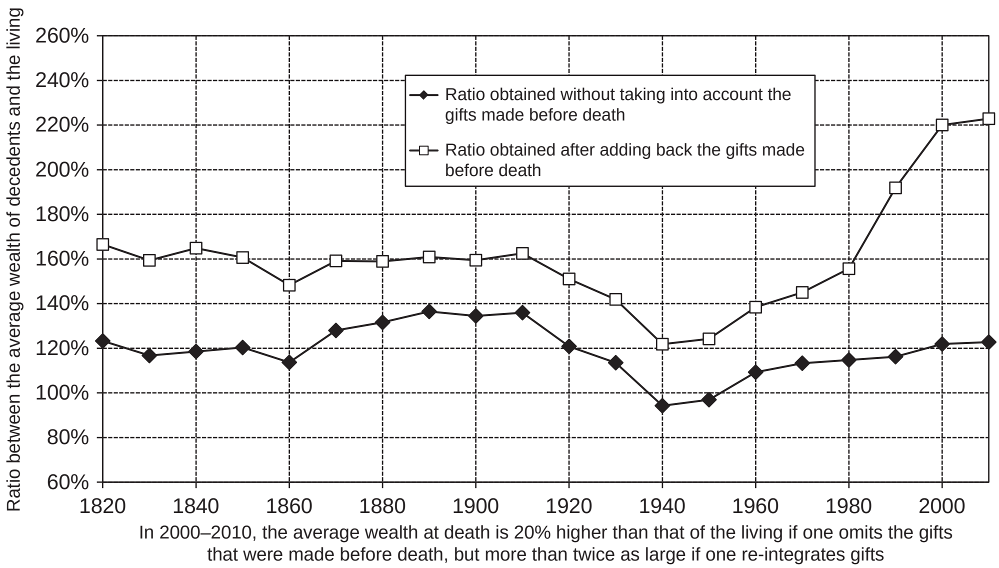


Figure 15.19 The ratio between average wealth at death and average wealth of the living, France 1820–2010.

In the recent decades, a very large part of the rise in $\mu_t^* = (1 + v_t) \cdot \mu_t$ comes from the rise in the gift–bequest ratio v_t, which used to be about 20% during most of the nineteenth to twentieth centuries, and has gradually risen to as much as 80% in recent decades (see Figure 15.19). That is, the gift flow is currently almost as large as the bequest flow.

Although there is still much uncertainty about the reasons behind the rise in gifts, the evidence suggests that it started before the introduction of new tax incentives for gifts in the 1990s–2000s, and has more to do with the growing awareness by wealthy parents that they will die old and that they ought to transmit part of their wealth inter-vivos if they want their children to fully benefit from it.

In any case, one should not underestimate the importance of gifts. In particular, one should not infer from a declining age–wealth profile at old ages or a relatively low relative wealth of decedents that inheritance is unimportant: this could simply reflect the fact that decedents have already given away a large part of their wealth.

15.4.2.2 The Inheritance Stock-Aggregate Wealth Ratio φ_t

How do the annual inheritance flows transmit into cumulated inheritance stocks? Given the data limitations we face, we show on Figure 15.20 two alternative estimates for the share φ_t of total inherited wealth in aggregate French wealth between 1850 and 2010. According to both measures, there is again a clear U-shaped pattern. The share of inherited wealth φ_t was as large as 80–90% of aggregate wealth in 1850–1910, down to as little as 35–45% around 1970, and back up to 65–75% by 2010.

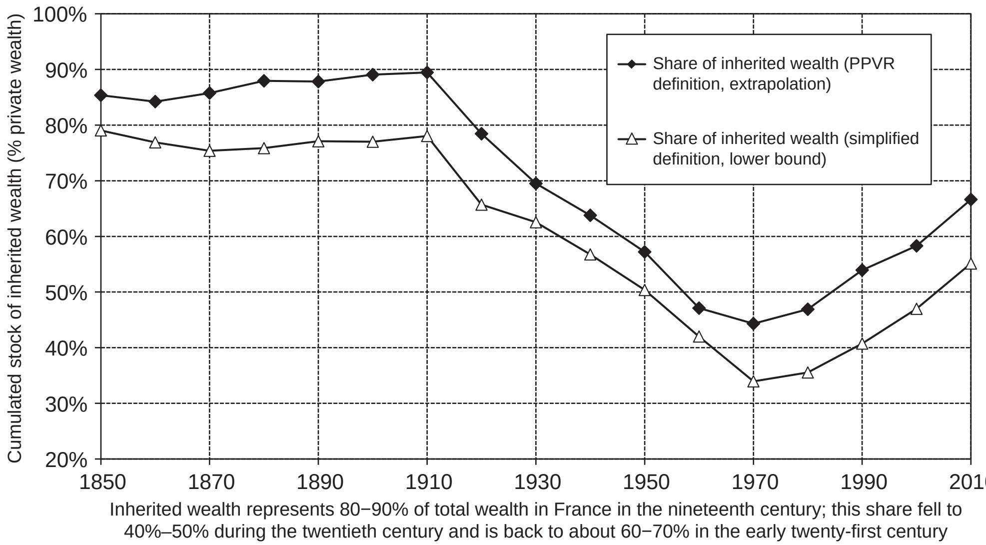


Figure 15.20 The cumulated stock of inherited wealth as a fraction of aggregate private wealth, France 1850–2010.

The higher series, which we see as the most reliable, was obtained by applying the micro-based PPVR definition (see Section 15.4.1.4). The limitation here is that the set of micro data on wealth over two generations that has been collected in French historical archives is more complete for Paris than for the rest of France (see Piketty et al., 2006, 2013). For years with missing data for the rest of France, the estimates reported on Figure 15.20 were extrapolated on the basis of the Parisian data. Ongoing data collection suggests that the final estimates will not be too different from the approximate estimates reported here.

The lower series, which we see as a lower bound, comes from the simplified definition based on the comparison of inheritance and saving flows (see Section 15.4.1.5).[26] The key advantage of this simplified definition is that it requires much less data: it can readily be computed from the inheritance flow series b_{yt} that were reported above. It delivers estimates of the inheritance share φ_t that are always somewhat below the micro-based estimates, with a gap that appears to be approximately constant. The gap seems to be due to the fact that the simplified definition attributes too much saving to pure labor earners with little inheritance.

In both series, the share φ_t of total inherited wealth in aggregate wealth reaches its lowest historical point in the 1970s, whereas the inheritance flow b_{yt} reaches its lowest point in the immediate aftermath of World War II. The reason is that the stock of

[26] The series was computed as $\varphi = b_y/(b_y + (1-\alpha)\cdot s)$ using 30-year averages for saving rates, capital shares, and inheritance flows.

inherited wealth comes from cumulating the inheritance flows of the previous decades—hence the time lag.

15.4.3 Evidence from Other Countries

What do we know about the importance of inheritance in countries other than France? A recent wave of research attempts to construct estimates of the inheritance flow–national income ratio b_{yt} in a number of European countries. The series constructed by Atkinson (2013) for Britain and Schinke (2013) for Germany show that b_{yt} has also followed a U-shaped pattern in these two countries over the past century (see Figure 15.21). Data limitations, however, make it difficult at this stage to make precise comparisons between countries.

For Britain, the inheritance flow b_{yt} of the late nineteenth to early twentieth centuries seems to be similar to that of France, namely about 20–25% of national income. The flow then falls following the 1914–1945 shocks, albeit less spectacularly than in France, and recovers in recent decades. Karagiannaki (2011), in a study of inheritance in the United Kingdom from 1984 to 2005, also finds a marked increase in that period. The rebound, however, seems to be less strong in Britain than in France, so that the inheritance flow appears smaller than in France today. We do not know yet whether this finding is robust. Available British series are pure "fiscal flow" series (as opposed to French series, for which we have both an "economic" and a "fiscal" estimate). As pointed out by Atkinson (2013), the main reason for the weaker British rebound in recent decades is that the gift–bequest ratio v_t has not increased at all according to fiscal data (v_t has remained relatively flat at a

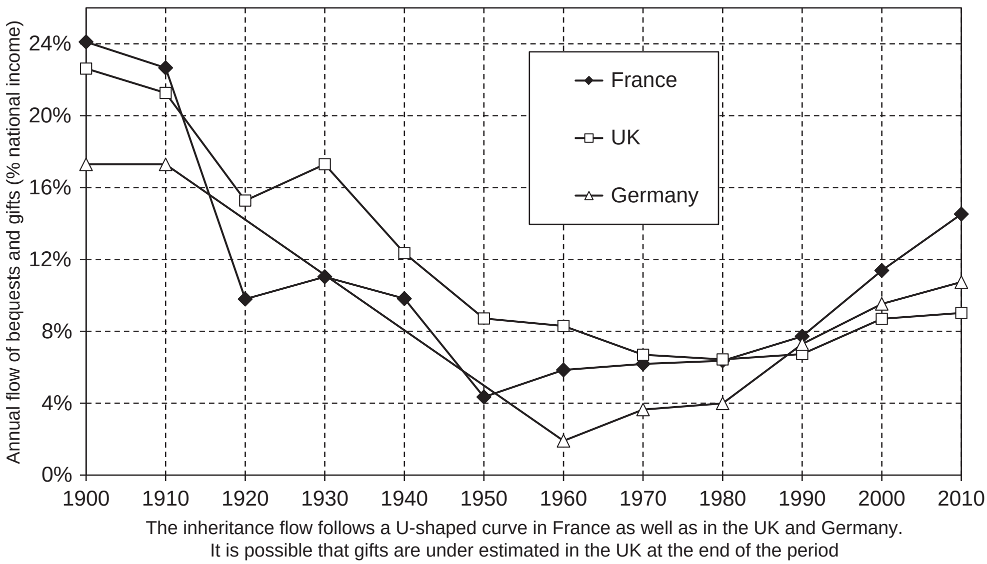
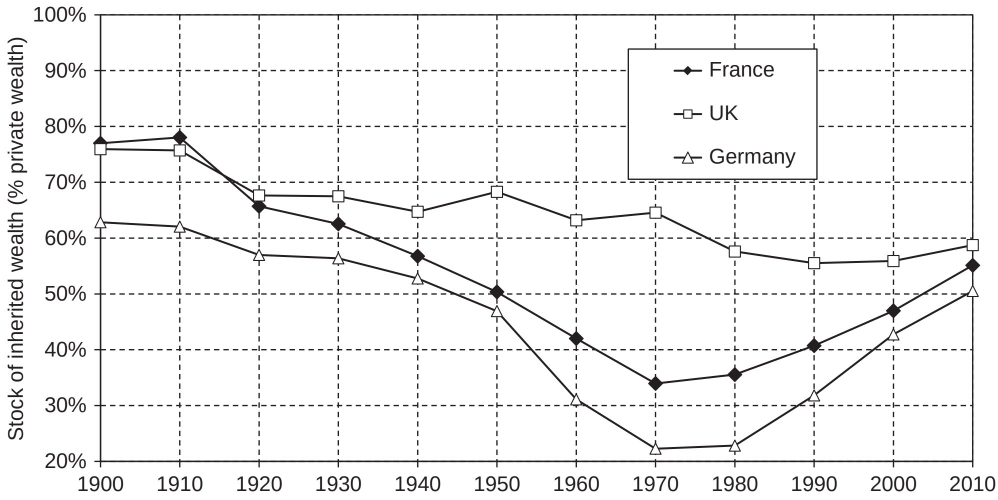


Figure 15.21 The inheritance flow in Europe, 1900–2010.



The inheritance share in aggregate wealth accumulation follows a U-shaped curve in France and Germany (and to a more limited extent in the UK and Germany). It is possible that gifts are underestimated in the UK at the end of the period

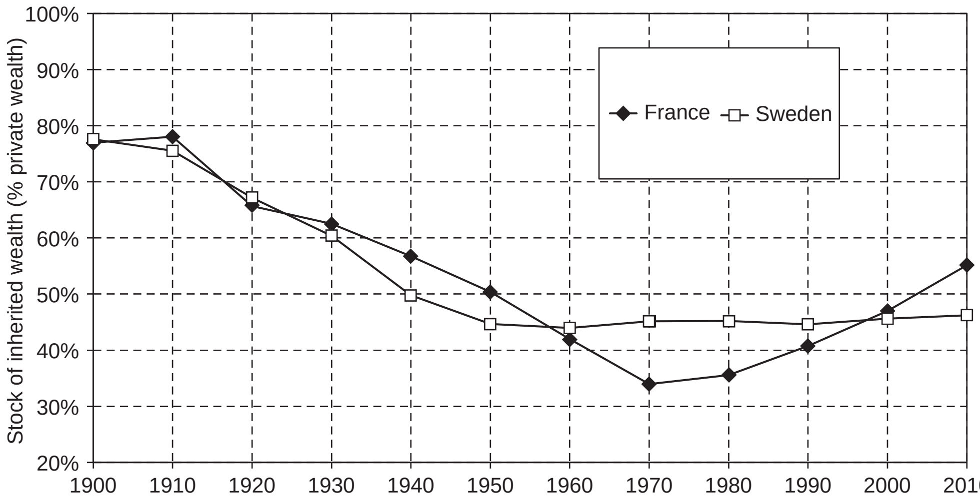
Figure 15.22 The inheritance stock in Europe, 1900–2010.

low level, around 10–20%). According to Atkinson, this could be due to substantial underreporting of gifts to tax authorities.

Germany also exhibits a U-shaped pattern of inheritance flow b_{yt} that seems to be broadly as sharp as in France. In particular, just as in France, the strong German rebound in recent decades comes with a large rise in the gift–bequest ratio v_t during the 1990s–2000s (v_t is above 50–60% in the 2000s). The overall levels of b_{yt} are generally lower in Germany than in France, which given the lower aggregate wealth–income ratio β_t is not surprising. Should we compare the rates of wealth transmission (i.e., $b_{wt} = b_{yt}/\beta_t$), then the levels would be roughly the same in both countries in 2000–2010.

We report on Figure 15.22 the corresponding estimates for the share φ_t of total inherited wealth in aggregate wealth, using the simplified definition $\varphi = b_y/(b_y + (1-\alpha)s)$. For Germany, the inheritance share φ_t appears to be generally smaller than in France. In particular, it reaches very low levels in the 1960s–1970s, owing to the extremely low inheritance flows in Germany in the immediate postwar period, and to large saving rates. In recent decades, the German φ_t has been rising fast and seems to catch up with France's. In the United Kingdom, the inheritance share φ_t apparently never fell to the low levels observed in France and Germany in the 1950s, and seems to be always higher than on the Continent. The reason, for the recent period, is that the United Kingdom has had relatively low saving rates since the 1970s.[27]

[27] In effect, British saving rates in recent decades are insufficient to explain the large rise in the aggregate wealth–income ratio, which can only be accounted for by large capital gain (Piketty and Zucman, 2014). The simplified definition of φ_t based on the comparison between inheritance and saving flows assumes the same capital gains for inherited and self-made wealth.



The inheritance share in aggregate wealth accumulation follows a broadly similar pattern in Sweden and France,although in recent decades the Swedish inheritance stock increased relatively little, as the private saving rate increased

Figure 15.23 The inheritance stock in France and Sweden, 1900–2010.

Recent historical research suggests that inheritance flows have also followed U-shaped patterns in Sweden (see Ohlsson et al., 2013). Here b_{yt} appears to be smaller than in France, but this again seems largely due to lower β_t ratios. When we look at the implied b_{wt} and φ_t ratios, which in a way are the most meaningful ratios to study, then both the levels and shape are relatively similar across European countries. As shown by Figure 15.23, the share of inherited wealth followed the same evolution in Sweden and France in the twentieth century (the main difference is that it seems to have increased a bit less in Sweden than in France in recent decades, because of a rise in the private saving rate). We stress again, however, that a lot more data needs to be collected—and to some extent is currently being collected—on the historical evolution of inheritance before we can make proper international comparisons.

Prior to the recent inheritance flow estimates surveyed above, a first wave of research, surveyed by Davies and Shorrocks (1999), mostly focused on the United States, with conflicting results—the famous Modigliani–Kotlikoff–Summers controversy. More recently, Kopczuk and Edlund (2009) observe that in estate tax data, the share of women among the very wealthy in the United States peaked in the late 1960s (at nearly one-half) and then declined to about one-third. They argue that this pattern reflects changes in the importance of inheritance, as women are less likely to be entrepreneurs. Wolff and Gittleman (2013) analyze Survey of Consumer Finances (SCF) data and find little evidence of a rise in inheritances since the late 1980s. Looking at *Forbes*' data, Kaplan and Rauh (2013) find that Americans in the Forbes 400 are less likely to have inherited their wealth today than in the 1980s. It is unclear, however, whether this result reflects a true economic phenomenon or illustrates the limits of *Forbes* and other wealth rankings.

Inherited wealth holdings are probably harder to spot than self-made wealth, first because inheritors' portfolios tend to be more diversified, and also because inheritors may not like to be in the press, while entrepreneurs usually enjoy it and do not attempt to dissimulate their wealth nearly as much. The conclusions about the relative importance of inherited versus self-made wealth obtained by analyzing *Forbe*s list data may thus be relatively fragile.

In the end, there remain important uncertainties about the historical evolution of inheritance in the United States. There are reasons to believe that inheritance has historically been less important in the United States than in Europe, because population growth has been much larger (more on this below). It is unclear whether this still applies today, however. Given the relatively low U.S. saving rates in recent decades, it is possible that even moderate inheritance flows imply a relatively large share φ_t of total inherited wealth in aggregate wealth (at least according to the simplified definition of φ based on the comparison between b_y and s).

One difficulty is that U.S. fiscal data on bequests and gifts are relatively low quality (in particular because the federal estate tax only covers few decedents; in 2012 only about 1 decedent out of 1000 pays the estate tax). One can use survey data (e.g., from the SCF) to estimate the relative wealth of decedents μ_t and compute the economic inheritance flow $b_{yt} = (1 + v_t) \cdot \mu_t \cdot \beta_t$. One key problem is that one needs to find ways to estimate the gift–bequest ratio v_t, which is not easy to do in the absence of high-quality fiscal data. Self-reported retrospective data on bequest and gift receipts usually suffer from large downward biases and should be treated with caution. In countries where there exists exhaustive administrative data on bequests and gifts (such as France, and to some extent Germany), survey-based self-reported flows appear to be less than 50% of fiscal flows. This may contribute to explain the low level of inheritance receipts found by Wolff and Gittleman (2013).[28]

15.5. ACCOUNTING FOR THE EVIDENCE: MODELS AND PREDICTIONS

15.5.1 Shocks Versus Steady States

How can we account for the historical evidence on the evolution of the aggregate wealth–income ratio, the concentration of wealth, and the share of inherited wealth? In this section, we describe the theoretical models that have been developed to address this question. While we still lack a comprehensive model able to rigorously and

[28] One additional challenge in this study is that inherited assets are generally valued using asset prices at the time the assets were transmitted: no capital gain is included—which probably contributes to a relatively low estimated inheritance share in total U.S. wealth (about 20%, just like in Modigiani's estimates). A comparison between inheritance flows and saving flows (using the simplified formula) would likely lead to more balanced results.

quantitatively assess the various effects at play, the literature makes it possible to highlight some of the key forces.

We are primarily concerned here about the determinants of long-run steady states. In practice, as should be clear from the historical series presented above, real-world economies often face major shocks and changes in fundamental parameters, so that we observe large deviations from steady states. In particular, the large decline in the aggregate wealth–income ratios β_t between 1910 and 1950 is due to the shocks induced by the two World Wars. By using detailed series on saving flows and war destructions, one can estimate the relative importance of the various factors at play (Piketty and Zucman, 2014). In the case of France and Germany, three factors of comparable magnitude each account for approximately one-third of the total 1910–1950 fall of β_t: insufficient national savings (a large part of private saving was absorbed by public deficits); war destructions; and the fall of relative assets prices (real estate and equity prices were historically low in 1950–1960, partly due to policies such as rent control and nationalization). In the case of Britain, war destructions were relatively minor, and the other two factors each account for about half of the fall in the ratio of wealth to income (war-induced public deficits were particularly large).[29]

In thinking about the future, is the concept of a steady state a relevant point of reference? Historical evidence suggests that it is. Whereas the dynamics of wealth and inequality has been chaotic in the twentieth century, eighteenth- and nineteenth-century United Kingdom and France can certainly be analyzed as being in a steady state characterized by low-growth, high wealth–income ratios, high levels of wealth concentration, and inheritance flows. This is true despite the fact that there were huge changes in the nature of wealth and of economic activity (from agriculture to industry).[30] The shocks of the twentieth century put an end to this steady state, and it seems justified to ask: if countries are to converge to a new steady state in the twenty-first century (that is, if the shocks of the twentieth century do not happen again), which long-term ratios will they reach?

We show that over a wide range of models, the long-run magnitude and concentration of wealth and inheritance are a decreasing function of g and an increasing function of $\bar{r}$, where g is the economy's growth rate and $\bar{r}$ is the net-of-tax rate of return to wealth. That is, under plausible assumptions, both the wealth–income ratio and the concentration of wealth tend to take higher steady-state values when the long-run growth rate is lower and when the net-of-tax rate of return is higher. In particular, a higher $\bar{r}-g$ tends to magnify steady-state wealth inequalities. Although there does not exist yet any

[29] For detailed decompositions of private and national wealth accumulation over the various subperiods, see Piketty and Zucman (2014).

[30] In particular, private wealth/income ratios and inheritance flows seemed quite stable in nineteenth-century France (with perhaps a slight upward trend at the end of the century), despite major structural economic changes. This suggests that although the importance of inheritance and wealth may rise and fall in response to the waves of innovation, a steady-state analysis is a fruitful perspective.

rigorous calibrations of these theoretical models, we argue that these predictions are broadly consistent with both the time-series and cross-country evidence. These findings also suggest that the current trends toward rising wealth–income ratios and wealth inequality might continue during the twenty-first century, both because of population and productivity-growth slowdown, and because of rising international competition to attract capital.

15.5.2 The Steady-State Wealth–Income Ratio: $\beta = s/g$

The most useful steady-state formula to analyze the long-run evolution of wealth–income and capital–output ratios is the Harrod–Domar–Solow steady-state formula:

$$\beta_t \to \beta = s/g.$$

With s = long-run (net-of-depreciation) saving rate, g = long-run growth rate.[31]

The steady-state formula $\beta = s/g$ is a pure accounting equation. By definition, it holds in the steady state of any micro-founded, one-good model of capital accumulation, independently of the exact nature of saving motives. It simply comes from the wealth-accumulation equation $W_{t+1} = W_t + S_t$, which can be rewritten in terms of wealth–income ratio $\beta_t = W_t/Y_t$:

$$\beta_{t+1} = \frac{\beta_t + s_t}{1 + g_t}$$

With $1 + g_t = Y_{t+1}/Y_t$ = growth rate of national income, $s_t = S_t/Y_t$ = net saving rate.

It follows immediately that if $s_t \to s$ and $g_t \to g$, then $\beta_t \to \beta = s/g$.

The Harrod–Domar–Solow says something trivial but important in a low-growth economy, the sum of capital accumulated in the past can become very large, as long as the saving rate remains sizable.

For instance, if the long-run saving rate is $s = 10\%$, and if the economy permanently grows at rate $g = 2\%$, then in the long run the wealth–income ratio has to be equal to $\beta = 500\%$, because it is the only ratio such that wealth rises at the same rate as income: $s/\beta = 2\% = g$. If the long-run growth rate declines to $g = 1\%$, and the economy keeps saving at rate $s = 10\%$, then the long-run wealth–income ratio will be equal to $\beta = 1000\%$.

In the long run, output growth g is the sum of productivity and population growth. In the standard one-good growth model, output is given by $Y_t = F(K_t, L_t)$, where K_t is non-human capital input and L_t is human labor input (i.e., efficient labor supply). L_t can be written as the product of raw-labor supply N_t and labor productivity parameter h_t. That is, $L_t = N_t \cdot h_t$, with $N_t = N_0 \cdot (1 + n)^t$ (n is the population growth rate) and $h_t = h_0 \cdot (1 + h)^t$ (h is

[31] When one uses gross-of-depreciation saving rates rather than net rates, the steady-state formula writes $\beta = s/(g + \delta)$ with s the gross saving rate, and δ the depreciation rate expressed as a proportion of the wealth stock.

the productivity growth rate). The economy's long-run growth rate g is given by the growth rate of L_t. Therefore it is equal to $1+g=(1+n)\cdot(1+h)$, i.e., $g\approx n+h$.[32] The long-run g depends both on demographic parameters (in particular, fertility rates) and on productivity-enhancing activities (in particular, the pace of innovation).

The long-run saving rate s also depends on many forces: s captures the strength of the various psychological and economic motives for saving and wealth accumulation (dynastic, life cycle, precautionary, prestige, taste for bequests, etc.). The motives and tastes for saving vary a lot across individuals and potentially across countries. Whether savings come primarily from a life cycle or a bequest motive, the $\beta=s/g$ formula will hold in steady state. In case saving is exogenous (as in the Solow model), the long-run wealth–income ratio will obviously be a decreasing function of the income growth rate g. This conclusion, however, is also true in a broad class of micro-founded, general equilibrium models of capital accumulation in which s can be endogenous and can depend on g. That is the case, in particular, in the infinite-horizon, dynastic model (in which s is determined by the rate of time preference and the concavity of the utility function), in "bequest-in-the-utility-function" models (in which the long-run saving rate s is determined by the strength of the bequest or wealth taste), and in most endogenous growth models (see box below). In all cases, for given preference parameters, the long-run $\beta=s/g$ tends to be higher when the growth rate is lower. A growth slowdown—coming from a decrease in population or productivity growth—tends to lead to higher capital–output and wealth–income ratios.

Box: The steady-state wealth–income ratio in macro models
Dynastic Model

Assume that output is given by $Y_t=F(K_t, L_t)$, where K_t is the capital stock and L_t is efficient labor and grows exogenously at rate g. Output is either consumed or added to the capital stock. We assume a closed economy, so the wealth–income ratio is the same as the capital–output ratio. In the infinite-horizon, dynastic model, each dynasty maximizes

$$V=\int_{t\geq s} e^{-\theta t}U(c_t)$$

where θ is the rate of time preference and $U(c_t)=c^{1-\gamma}(1-\gamma)$ is a standard utility function with a constant intertemporal elasticity of substitution equal to $1/\gamma$. This elasticity of substitution is often found to be small, typically between 0.2 and 0.5, and is in any case smaller than one. Therefore γ is typically bigger than one.

Continued

[32] To obtain the exact equality $g=n+h$, one needs to use instantaneous (continuous-time) growth rates rather than annual (discrete-time) growth rates. That is, with $N_t=N_0\cdot e^{nt}$ (with n=population growth rate) and $h_t=h_0\cdot e^{ht}$, we have $L_t=N_t\cdot h_t=L_0\cdot e^{gt}$, with $g=n+h$.

The first-order condition describing the optimal consumption path of each dynasty is: $\mathrm{d}c_t/\mathrm{d}t = (r - \theta) \cdot c_t/\gamma$, i.e., utility-maximizing agents want their consumption path to grow at rate $g_c = (r - \theta)/\gamma$. This is a steady state if and only if $g_c = g$, i.e., $r = \theta + \gamma g$, what is known as the modified Golden Rule of capital accumulation. The long-run rate of return $r = \theta + \gamma g$ is entirely determined by preference parameters and the growth rate and is larger than g.

The steady-state saving rate is equal to $s = \alpha \cdot g/r = \alpha \cdot g/(\theta + \gamma g)$, where $\alpha = r \cdot \beta$ is the capital share. Intuitively, a fraction g/r of capital income is saved in the long run, so that dynastic wealth grows at the same rate g as national income. The saving rate $s = s(g)$ is an increasing function of the growth rate, but rises less fast than g, so that the steady-state wealth–income ratio $\beta = s/g$ is a decreasing function of the growth rate.

For instance, with a Cobb–Douglas production function (in which case the capital share is entirely set by technology and is constantly equal to α), the wealth–income ratio is given by $\beta = \alpha/r = \alpha/(\theta + \gamma \cdot g)$ and takes its maximum value $\overline{\beta} = \alpha/\theta$ for $g = 0$.

One unrealistic feature of the dynastic model is that it assumes an infinite long-run elasticity of capital supply with respect to the net-of-tax rate of return, which mechanically entails extreme consequences for optimal capital tax policy (namely, zero tax). The "bequest-in-the-utility-function" model provides a less extreme and more flexible conceptual framework in order to analyze the wealth accumulation process.

Wealth-in-the-Utility-Function Model

Consider a dynamic economy with a discrete set of generations 0, 1, …, t, …, zero population growth, and exogenous labor productivity growth at rate $g > 0$. Each generation has measure $N_t = N$, lives one period, and is replaced by the next generation. Each individual living in generation t receives bequest $b_t = w_t \geq 0$ from generation $t-1$ at the beginning of period t, inelastically supplies one unit of labor during his lifetime (so that labor supply $L_t = N_t = N$), and earns labor income y_{Lt}. At the end of period, he then splits lifetime resources (the sum of labor income and capitalized bequests received) into consumption c_t and bequests left $b_{t+1} = w_{t+1} \geq 0$, according to the following budget constraint:

$$c_t + b_{t+1} \leq y_t = y_{Lt} + (1 + r_t) b_t$$

The simplest case is when the utility function is defined directly over consumption c_t and the increase in wealth $\Delta w_t = w_{t+1} - w_t$ and takes a simple Cobb–Douglas form: $V(c, \Delta w) = c^{1-s} \Delta w^s$. (Intuitively, this corresponds to a form of "moral" preferences where individuals feel that they cannot possibly leave less wealth to their children than what they have received from their parents, and derive utility from the increase in wealth, maybe because this is a signal of their ability or virtue.) Utility maximization then leads to a fixed saving rate: $w_{t+1} = w_t + s y_t$. By multiplying per capita values by population $N_t = N$ we have the same linear transition equation at the aggregate level: $W_{t+1} = W_t + s Y_t$. The long-run wealth–income ratio is given by $\beta_t \rightarrow \beta = s/g$. It depends on the strength of the bequest motive and on the rate of productivity growth.

With other functional forms for the utility function, e.g., with $V = V(c, w)$, or with heterogenous labor productivities or saving tastes across individuals, one simply needs to replace the parameter s by the properly defined average wealth or bequest taste parameter. For instance, with $V(c, w) = c^{1-s} w^s$, utility maximization leads to $w_{t+1} = s \cdot (w_t + y_t)$ and

Continued

$\beta_t \rightarrow \beta = s/(g+1-s) = \tilde{s}/g$, with $\tilde{s} = s(1+\beta) - \beta$ the conventional saving rate (i.e., defined relative to income). See Section 15.5.4.1 for a simple application of this model to the analysis of the steady-state distribution of wealth.

Endogenous Growth Models

In endogenous growth models with imperfect international capital flows, the growth rate might rise with the saving rate, but it will usually rise less than proportionally. It is only in what is known as the AK closed-economy model that the growth rate rises proportionally with the saving rate. To see this, assume zero population growth ($n=0$) and a Cobb–Douglas production function $Y = K^{\alpha} \cdot (A_L \cdot L)^{1-\alpha}$. Further assume that the productivity parameter is endogenously determined by an economy-wide capital accumulation externality, such that $A_L = A_0 \cdot K$. Then we have $Y = A \cdot K$, with $A = (A_0 \cdot L_0)^{1-\alpha}$. For a given saving rate $s > 0$, the growth rate is given by $g = g(s) = s \cdot A$. The growth rate rises proportionally with the saving rate, so that the wealth–income ratio is entirely set by technology: $\beta = s/g = 1/A$ is a constant.

In more general endogenous growth models, the rate of productivity growth depends not only on the pace of capital accumulation, but also—and probably more importantly—on the intensity of innovation activities, the importance of education spendings, the position on the international technological frontier, and a myriad of other policies and institutions, so that the resulting growth rate rises less than proportionally with the saving rate.

The slowdown of income growth is the central force explaining the rise of wealth–income ratios in rich countries over the 1970–2010 period, particularly in Europe and Japan, where population growth has slowed markedly (and where saving rates are still high relative to the United States). As Piketty and Zucman (2014) show, the cumulation of saving flows explains the 1970–2010 evolution of β in the main rich countries relatively well. An additional explanatory factor over this time period is the gradual recovery of relative asset prices. In the very long run, however, relative asset-price movements tend to compensate each other, and the one-good capital accumulation model seems to do a good job at explaining the evolution of wealth–income ratios.

It is worth stressing that the $\beta = s/g$ formula works both in closed-economy and open-economy settings. The only difference is that wealth–income and capital–output ratios are the same in closed-economy settings but can differ in open-economy environments.

In the closed-economy case, private wealth is equal to domestic capital: $W_t = K_t$.[33] National income Y_t is equal to domestic output $Y_{dt} = F(K_t, L_t)$. Saving is equal to domestic investment, and the private wealth–national income ratio $\beta_t = W_t/Y_t$ is the same as the domestic capital–output ratio $\beta_{kt} = K_t/Y_{dt}$.

In the open economy case, countries with higher saving rates $s_a > s_b$ accumulate higher wealth ratios $\beta_a = s_a/g > \beta_b = s_b/g$ and invest some their wealth in countries with

[33] For simplicity we assume away government wealth and saving.

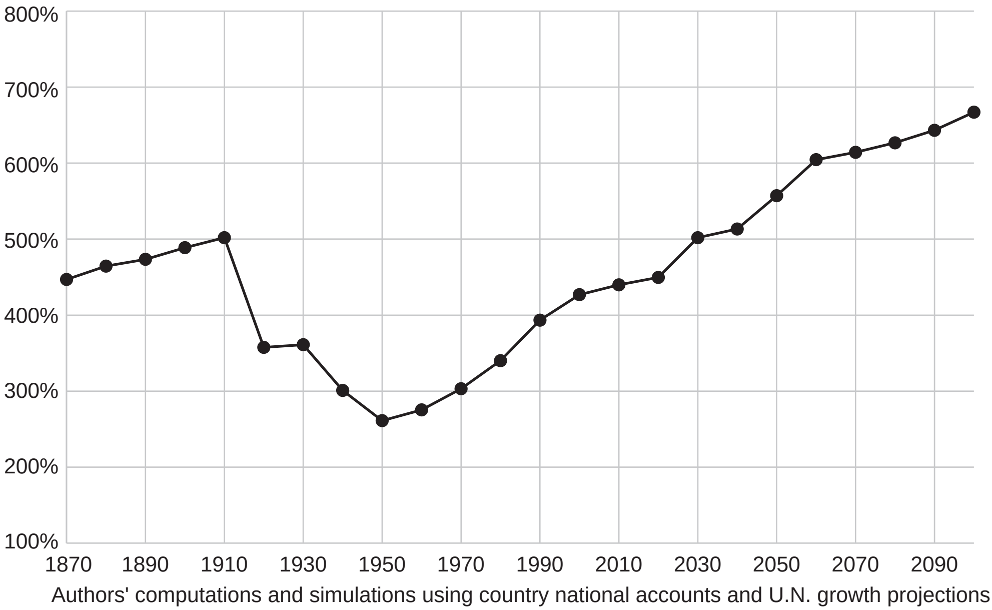


Figure 15.24 World wealth/national income ratio, 1870–2100.

lower saving rates, so that the capital–output ratio is the same everywhere (assuming perfect capital mobility). Noting N_a and N_b the population of countries a and b, $N=N_a+N_b$ the world population, $Y=Y_a+Y_b$ the world output, and $s=(s_a \cdot Y_a+s_b \cdot Y_b)/Y$ the world saving rate, and assuming that each country's effective labor supply is proportional to population and grows at rate g, then the long-run wealth–income and capital–output ratio at the world level will be equal to $\beta=s/g$. With perfect capital mobility, each country will operate with the same capital–output ratio $\beta=s/g$. Country a with wealth $\beta_a>\beta$ will invest its extra wealth $\beta_a-\beta$ in country b with wealth $\beta_b<\beta$. Both countries have the same per capita output $y=Y/N$, but country a has a permanently higher per capita national income $y_a=y+r\cdot(\beta_a-\beta)>y$, while country b has a permanently lower per capita national income $y_b=y-r\cdot(\beta-\beta_b)<y$. In the case of Britain and France at the eve of World War I, the net foreign wealth position $\beta_a-\beta$ was of the order of 100–200%, the return on net foreign assets was about $r=5\%$, so that national income was about 5–10% larger than domestic output.

At the world level, wealth–income and capital–output ratios always coincide (by definition). The long-run ratio is governed by the steady-state condition $\beta=s/g$. In the very long run, if the growth rate slows down at the global level (in particular due to the possible stabilization of world population), then the global β might rise. We report on Figure 15.24 one possible evolution of the world wealth–income ratio in the twenty-first century, assuming that the world-income growth rate stabilizes at about 1.5% and world saving rate at about 12%. Under these (arguably specific and uncertain) assumptions, the world β would rise to about 700–800% by the end of the twenty-first century.

15.5.3 The Steady-State Capital Share: $\alpha = r \cdot \beta = a \cdot \beta^{\frac{\sigma-1}{\sigma}}$

How does the evolution of the capital–income ratio β relate to the evolution of the capital share $\alpha_t = r_t \cdot \beta_t$ (where r_t is the average rate of return)? All depends on whether the capital–labor elasticity of substitution σ is larger or smaller than one.

Take a CES production function $Y = F(K, L) = \left(a \cdot K^{\frac{\sigma-1}{\sigma}} + (1-a) \cdot L^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma-1}{\sigma}}$. The rate of return is given by $r = F_K = a\beta^{-1/\sigma}$ (with $\beta = K/Y$), and the capital share is given by $\alpha = r \cdot \beta = a \cdot \beta^{\frac{\sigma-1}{\sigma}}$. If $\sigma > 1$, then as β_t rises, the fall of the marginal product of capital r_t is smaller than the rise of β_t, so that the capital share $\alpha_t = r_t \cdot \beta_t$ is an increasing function of β_t. Conversely, if $\sigma < 1$, the fall of r_t is bigger than the rise of β_t, so that the capital share is a decreasing function of β_t.[34]

As $\sigma \to \infty$, the production function becomes linear, that is, the return to capital is independent of the quantity of capital: this is like a robot economy where capital can produce output on its own. Conversely, as $\sigma \to 0$, the production function becomes putty clay, that is, the return to capital falls to zero if the quantity of capital is slightly above the fixed-proportion technology.

A special case is when the capital–labor elasticity of substitution σ is exactly equal to one: changes in r and in β exactly compensate each other so that the capital share is constant. This is the Cobb–Douglas case $F(K, L) = K^{\alpha} L^{1-\alpha}$. The capital share is entirely set by technology: $\alpha_t = r_t \cdot \beta_t = \alpha$. A higher capital–output ratio β_t is exactly compensated by a lower capital return $r_t = \alpha/\beta_t$, so that the product of the two is constant.

There is a large literature trying to estimate the elasticity of substitution between labor and capital, reviewed in Antras (2004) and Chirinko (2008); see also Karabarbounis and Neiman (2014). The range of estimates is wide. Historical evidence suggests that the elasticity of substitution σ may have risen over the development process. In the eighteenth to nineteenth centuries, it is likely that σ was less than one, particularly in the agricultural sector. An elasticity less than one would explain why countries with large quantities of land (e.g., the United States) had lower aggregate land values than countries with little land (the Old World). Indeed, when $\sigma < 1$, price effects dominate volume effects: when land is very abundant, the price of land is extremely low, and the product of the two is small. An elasticity less than 1 is exactly what one would accept in an economy in which

[34] Because we include all forms of capital assets into our aggregate capital concept K, the aggregate elasticity of substitution σ should be interpreted as resulting from both supply forces (producers shift between technologies with different capital intensities) and demand forces (consumers shift between goods and services with different capital intensities, including housing services versus other goods and services).

capital takes essentially one form only (land), as in the eighteenth and early nineteenth centuries. When there is too much of the single capital good, it becomes almost useless.

Conversely, in the twentieth century, capital shares α have tended to move in the same direction as capital–income ratios β. This fact suggests that the elasticity of substitution σ has been larger than one. Since the mid-1970s, in particular, we do observe a significant rise of capital shares α_t in rich countries (Figure 15.25). Admittedly, the rise in capital shares α_t was less marked than the rise of capital–income ratios β_t—in other words, the average return to wealth $r_t = \alpha_t/\beta_t$ has declined (Figure 15.26). But this decline is

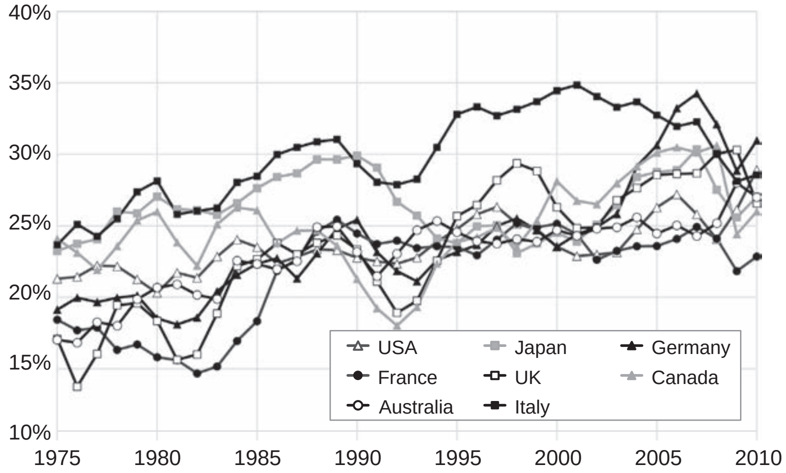


Figure 15.25 Capital shares in factor-price national income, 1975–2010.

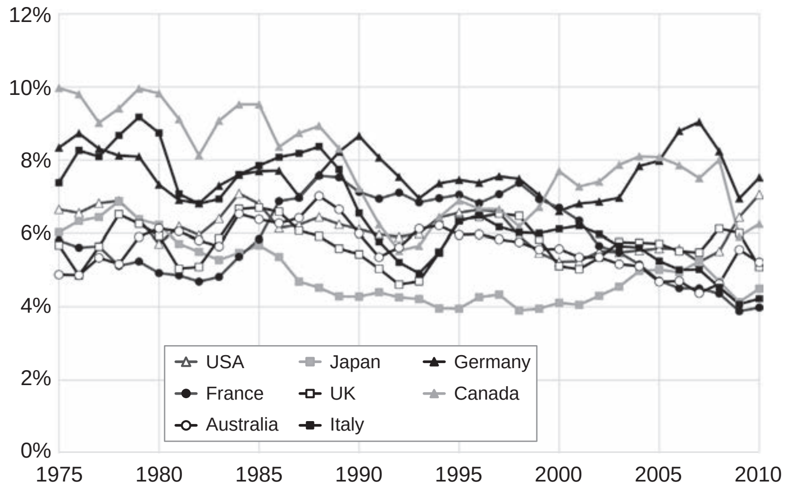


Figure 15.26 Average return on private wealth, 1975–2010.

exactly what one should expect in any economic model: when there is more capital, the rate of return to capital must go down. The interesting question is whether the average return r_t declines less or more than β_t increases. The data gathered by Piketty and Zucman (2014) suggest that r_t has declined less, i.e., that the capital share has increased, consistent with an elasticity $\sigma > 1$. This result is intuitive: an elasticity larger than one is what one would expect in a sophisticated economy with different uses for capital (not only land, but also robots, housing, intangible capital, etc.). The elasticity might even increase with globalization, as it becomes easier to move different forms of capital across borders.

Importantly, the elasticity does not need to be hugely superior to one in order to account for the observed trends. With an elasticity σ around 1.2–1.6, a doubling of capital–output ratio β can lead to a large rise in the capital share α. With large changes in β, one can obtain substantial movements in the capital share with a production function that is only moderately more flexible than the standard Cobb–Douglas function. For instance, with $\sigma = 1.5$, the capital share rises from $\alpha = 28\%$ to $\alpha = 36\%$ if the wealth–income ratio jumps from $\beta = 2.5$ to $\beta = 5$, which is roughly what has happened in rich countries since the 1970s. The capital share would reach $\alpha = 42\%$ in case further capital accumulation takes place and the wealth–income ratio attains $\beta = 8$. In case the production function becomes even more flexible over time (say, $\sigma = 1.8$), the capital share would then be as large as $\alpha = 53\%$.[35] The bottom line is that we certainly do not need to go all the way toward a robot economy ($\sigma = \infty$) in order to generate very large movements in the capital share.

15.5.4 The Steady-State Level of Wealth Concentration: $Ineq = Ineq\ (\bar{r} - g)$

The possibility that the capital–income ratio β—and maybe the capital share α—might rise to high levels entails very different welfare consequences depending on who owns capital. As we have seen in Section 15.3, wealth is always significantly more concentrated than income, but wealth has also become less concentrated since the nineteenth to early twentieth century, at least in Europe. The top 10% wealth holders used to own about 90% of aggregate wealth in Europe prior to World War I, whereas they currently own about 60–70% of aggregate wealth.

What model do we have to analyze the steady-state level of wealth concentration? There is a large literature devoted to this question. Early references include Champernowne (1953), Vaughan (1979), and Laitner (1979). Stiglitz (1969) is the first attempt to analyze the steady-state distribution of wealth in the neoclassical growth model. In his and similar models of wealth accumulation, there is at the same time both convergence of the macro-variables to their steady-state values and of the distribution of wealth to its steady-state form. Dynamic wealth-accumulation models with random

[35] With $a = 0.21$ and $\sigma = 1.5$, $\alpha = a \cdot \beta^{\frac{\sigma-1}{\sigma}}$ goes from 28% to 36% and 42% as β rises from 2.5 to 5 and 8. With $\sigma = 1.8$, α rises to 53% if $\beta = 8$.

idiosyncratic shocks have the additional property that a higher $\bar{r}-g$ differential (where $\bar{r}$ is the net-of-tax rate of return to wealth and g is the economy's growth rate) tends to magnify steady-state wealth inequalities. This is particularly easy to see in dynamic model with random multiplicative shocks, where the steady-state distribution of wealth has a Pareto shape, with a Pareto exponent that is directly determined by $\bar{r}-g$ (for a given structure of shocks).

15.5.4.1 An Illustrative Example with Closed-Form Formulas

To illustrate this point, consider the following model with discrete time $t=0, 1, 2, \ldots$. The model can be interpreted as an annual model (with each period lasting $H=1$ year), or a generational model (with each period lasting $H=30$ years), in which case saving tastes can be interpreted as bequest tastes. Suppose a stationary population $N_t=[0, 1]$ made of a continuum of agents of size one, so that aggregate and average variables are the same for wealth and national income: $W_t=w_t$ and $Y_t=y_t$. Effective labor input $L_t=N_t \cdot h_t=h_0 \cdot (1+g)^t$ grows at some exogenous, annual productivity rate g. Domestic output is given by some production function $Y_{dt}=F(K_t, L_t)$.

We suppose that each individual $i \in [0, 1]$ receives the same labor income $y_{Lti}=y_{Lt}$ and has the same annual rate of return $r_{ti}=r_t$. Each agent chooses c_{ti} and w_{t+1i} so as to maximize a utility function of the form $V(c_{ti}, w_{ti})=c_{ti}^{1-s_{ti}} w_{ti}^{s_{ti}}$, with wealth (or bequest) taste parameter s_{ti} and budget constraint $c_{ti}+w_{t+1i} \leq y_{Lt}+(1+r_t) \cdot w_{ti}$. Random shocks only come from idiosyncratic variations in the saving taste parameters s_{ti}, which are supposed to be drawn according to some i.i.d. random process with mean $s=E(s_{ti})<1$.[36]

With the simple Cobb–Douglas specification for the utility function, utility maximization implies that consumption c_{ti} is a fraction $1-s_{ti}$ of $y_{Lt}+(1+r_t) \cdot w_{ti}$, the total resources (income plus wealth) available at time t. Plugging this formula into the budget constraint, we have the following individual-level transition equation for wealth:

$$w_{t+1i}=s_{ti} \cdot [y_{Lt}+(1+r_t) \cdot w_{ti}] \quad (15.1)$$

At the aggregate level, since by definition national income is equal to $y_t=y_{Lt}+r_t \cdot w_t$, we have

$$w_{t+1}=s \cdot [y_{Lt}+(1+r_t) \cdot w_t]=s \cdot [y_t+w_t] \quad (15.2)$$

dividing by $y_{t+1} \approx (1+g) \cdot y_t$ and denoting $\alpha_t=r_t \cdot \beta_t$ the capital share and $(1-\alpha_t)=y_{Lt}/y_t$ the labor share, we have the following transition equation for the wealth–income ratio $\beta_t=w_t/y_t$:

[36] For a class of dynamic stochastic models with more general structures of preferences and shocks, see Piketty and Saez (2013).

$$\beta_{t+1} = s \cdot \frac{1-\alpha_t}{1+g} + s \cdot \frac{1+r_t}{1+g} \cdot \beta_t = \frac{s}{1+g} \cdot (1+\beta_t) \tag{15.3}$$

In the open-economy case, the world rate of return $r_t = r$ is given. From the above equation one can easily see that β_t converges toward a finite limit β if and only if

$$\omega = s \cdot \frac{1+r}{1+g} < 1$$

In case $\omega > 1$, then $\beta_t \to \infty$. In the long run, the economy is no longer a small open economy, and the world rate of return will have to fall so that $\omega < 1$.

In the closed-economy case, β_t always converges toward a finite limit, and the long-run rate of return r is equal to the marginal product of capital and depends negatively upon β. With a CES production function, for example, we have $r = F_K = \alpha \cdot \beta^{-1/\sigma}$ (see Section 15.5.3).

Setting $\beta_{t+1} = \beta_t$ in Equation (15.3), we obtain the steady-state wealth–income ratio:

$$\beta_t \to \beta = s/(g+1-s) = \tilde{s}/g$$

where $\tilde{s} = s(1+\beta) - \beta$ is the steady-state saving rate expressed as a fraction of national income.

Noting $z_{ti} = w_{ti}/w_t$ the normalized individual wealth, and dividing both sides of Equation (15.1) by $w_{t+1} \approx (1+g) \cdot w_t$, the individual-level transition equation for wealth can be rewritten as follows[37]:

$$z_{t+1i} = \frac{s_{ti}}{s} \cdot [(1-\omega) + \omega \cdot z_{ti}] \tag{15.4}$$

Standard convergence results (e.g., Hopehnayn and Prescott, 1992, Theorem 2, p. 1397) then imply that the distribution $\psi_t(z)$ of relative wealth will converge toward a unique steady-state distribution $\psi(z)$ with a Pareto shape and a Pareto exponent that depends on the variance of taste shocks s_{ti} and on the ω coefficient.

For instance, assume simple binomial taste shocks: $s_{ti} = s_0 = 0$ with probability $1-p$, and $s_{ti} = s_1 > 0$ with probability p (with $s = p \cdot s_1$ and $\omega < 1 < \omega/p$). The long-run distribution function $1 - \Psi_t(z) = \text{proba}(z_{ti} \geq z)$ will converge for high z toward

$$1 - \Phi(z) \approx \left(\frac{\lambda}{z}\right)^a,$$

with a constant term λ

[37] Note that $y_{Lt} = (1-\alpha) \cdot y_t$, where $\alpha = r \cdot \beta = r \cdot s/(1+g-s)$ is the long-run capital share. Note also that the individual-level transition equation given below holds only in the long run (i.e., when the aggregate wealth–income ratio has already converged).

$$\lambda = \frac{1-\omega}{\omega - p},$$

a Pareto coefficient a

$$a = \frac{\log(1/p)}{\log(\omega/p)} > 1, \tag{15.5}$$

and an inverted Pareto coefficient b

$$b = \frac{a}{a-1} = \frac{\log(1/p)}{\log(1/\omega)} > 1.$$

To see this, note that the long-run distribution with $\omega < 1 < \omega/p$ looks as follows: $z = 0$ with probability $1-p$, $z = \frac{1-\omega}{p}$ with probability $(1-p)\cdot p$, …, and $z = z_k = \frac{1-\omega}{\omega-p}\cdot\left[\left(\frac{\omega}{p}\right)^k - 1\right]$ with probability $(1-p)\cdot p^k$. As $k \to +\infty$, $z_k \approx \frac{1-\omega}{\omega-p}\cdot\left(\frac{\omega}{p}\right)^k$ The cumulated distribution is given by $1-\Phi(z_k) = \text{proba}(z \geq z_k) = \sum_{k' \geq k}(1-p)\cdot p^{k'} = p^k$. It follows that as $z \to +\infty$, $\log[1-\Phi(z)] \approx a\cdot[\log(\lambda) - \log(z)]$, i.e., $1-\Phi(z) \approx (\lambda/z)^a$. In case $\omega/p < 1$, then $z_k = \frac{1-\omega}{p-\omega}\cdot\left[1 - \left(\frac{\omega}{p}\right)^k\right]$ has a finite upper bound $z_1 = \frac{1-\omega}{p-\omega}$.[38]

As ω rises, a declines and b rises, which means that the steady-state distribution of wealth is more and more concentrated.[39] Intuitively, an increase in $\omega = s\cdot\frac{1+r}{1+g}$ means that the multiplicative wealth inequality effect becomes larger as compared to the equalizing labor income effect, so that steady-state wealth inequalities get amplified.

In the extreme case where $\omega \to 1^-$ (for given $p < \omega$), $a \to 1^+$ and $b \to +\infty$ (infinite inequality). That is, the multiplicative wealth inequality effect becomes infinite as compared to the equalizing labor-income effect. The same occurs as $p \to 0^+$ (for given $\omega > p$): an infinitely small group gets infinitely large random shocks.[40] Explosive wealth inequality paths can also occur in case the taste parameter s_{ti} is higher on average for individuals with high initial wealth.[41]

[38] See Piketty and Saez (2013, working paper version, pp. 51–52).

[39] A higher inverted Pareto coefficient b (or, equivalently, a lower Pareto coefficient a) implies a fatter upper tail of the distribution and higher inequality. On the historical evolution of Pareto coefficients, see Atkinson et al. (2011, pp. 13–14 and 50–58).

[40] In the binomial model, one can directly compute the "empirical" inverted Pareto coefficient $b' = \frac{E(z|z \geq z_k)}{z_k} \to \frac{1-p}{1-\omega}$ as $k \to +\infty$. Note that $b' \simeq b$ if $p, \omega \simeq 1$, but that the two coefficients generally differ because the true distribution is discrete, while the Pareto law approximation is continuous.

[41] Kuznets (1953) and Meade (1964) were particularly concerned about this potentially powerful unequalizing force.

15.5.4.2 Pareto Formulas in Multiplicative Random Shocks Models

More generally, one can show that all models with multiplicative random shocks in the wealth accumulation process give rise to distributions with Pareto upper tails, whether the shocks are binomial or multinomial, and whether they come from tastes or other factors. For instance, the shock can come from the rank of birth, such as in the primogeniture model of Stiglitz (1969),[42] or from the number of children (Cowell, 1998),[43] or from rates of return (Benhabib et al., 2011, 2013; Nirei, 2009). Whenever the transition equation for wealth can be rewritten so as take a multiplicative form

$$z_{t+1i} = \omega_{ti} \cdot z_{ti} + \varepsilon_{ti}$$

where ω_{ti} is an i.i.d. multiplicative shock with mean $\omega = E(\omega_{ti}) < 1$, and ε_{ti} an additive shock (possibly random), then the steady-state distribution has a Pareto upper tail with coefficient a, which must solve the following equation:

$$E\left(\omega_{ti}^{a}\right) = 1.$$

A special case is when $p \cdot (\omega/p)^a = 1$, that is $a = \log(1/p)/\log(\omega/p)$, the formula given in Equation (15.5) above. More generally, as long as $\omega_{ti} > 1$ with some positive probability, there exists a unique $a > 1$, so that $E(\omega_{ti}^{a}) = 1$. One can easily see that for a given average $\omega = E(\omega_{ti}) < 1$, $a \to 1$ (and thus wealth inequality tends to infinity) if the variance of shocks goes to infinity, and $a \to \infty$ if the variance goes to zero.

Which kind of shocks have mattered most in the historical dynamics of the distribution of wealth? Many different kinds of individual-level random shocks play an important role in practice, and it is difficult to estimate the relative importance of each of them. One robust conclusion, however, is that for a given variance of shocks, steady-state wealth concentration is always a rising function of $r - g$. That is, due to cumulative dynamic effects, relatively small changes in $r - g$ (say, from $r - g = 2\%$ per year to $r - g = 3\%$ per year) can make a huge difference in terms of long-run wealth inequality.

For instance, if we interpret each period of the discrete-time model described above as lasting H years (with $H = 30$ years = generation length), and if r and g denote instantaneous rates, then the multiplicative factor ω can be rewritten as

[42] With primogeniture (binomial shock), the formula is exactly the same as before. See, e.g., Atkinson and Harrison (1978, p. 213), who generalize the Stiglitz (1969) formula and get: $a = \log(1+n)/\log(1+sr)$, with s the saving rate out of capital income. This is the same formula as $a = \log(1/p)/\log(\omega/p)$: with population growth rate per generation $= 1+n$, the probability that a good shock occurs—namely, being the eldest son—is given by $p = 1/(1+n)$. Menchik (1980), however, provides evidence on estate division in the United States, showing that equal sharing is the rule.

[43] The Cowell result is more complicated because families with many children do not return to zero (unless infinite number of children), so there is no closed form formula for the Pareto coefficient a, which must solve the following equation: $\sum \frac{p_k \cdot k}{2}\left(\frac{2 \cdot \omega}{k}\right)^a = 1$, where p_k = fraction of parents who have k children, with $k = 1, 2, 3$, etc., and ω = average generational rate of wealth reproduction.

$$\omega = s \cdot \frac{1+R}{1+G} = s \cdot e^{(r-g)H}$$

with $1+R=e^{rH}$ the generational rate of return and $1+G=e^{gH}$ the generational growth rate. If $r-g$ rises from $r-g=2\%$ to $r-g=3\%$, then with $s=20\%$ and $H=30$ years, $\omega = s \cdot e^{(r-g)H}$ rises from $\omega=0.36$ to $\omega=0.49$. For a given binomial shock structure $p=10\%$, this implies that the resulting inverted Pareto coefficient $b=(\log(1/p))/(\log(1/\omega))$ shifts from $b=2.28$ to $b=3.25$. This corresponds to a shift from an economy with moderate wealth inequality (say, with a top 1% wealth share around 20–30%) to an economy with very high wealth inequality (say, with a top 1% wealth share around 50–60%).

Last, if we introduce taxation into the dynamic wealth accumulation model, then one naturally needs to replace r by the after-tax rate of return $\bar{r}=(1-\tau)\cdot r$, where τ is the equivalent comprehensive tax rate on capital income, including all taxes on both flows and stocks. That is, what matters for long-run wealth concentration is the differential $\bar{r}-g$ between the net-of-tax rate of return and the growth rate. This implies that differences in capital tax rates and tax progressivity over time and across countries can explain large differences in wealth concentration.[44]

15.5.4.3 On the Long-Run Evolution of $\bar{r}-g$

The fact that steady-state wealth inequality is a steeply increasing function of $\bar{r}-g$ can help explain some of the historical patterns analyzed in Section 15.3.

First, it is worth emphasizing that during most of history, the gap $\bar{r}-g$ was large, typically of the order of 4–5% per year. The reason is that growth rates were close to zero until the industrial revolution (typically less than 0.1–0.2% per year), while the rate of return to wealth was generally of the order of 4–5% per year, in particular for agricultural land, by far the most important asset.[45] We have plotted on Figure 15.27 the world GDP growth rates since Antiquity (computed from Maddison, 2010) and estimates of the average return to wealth (from Piketty, 2014). Tax rates were negligible prior to the twentieth century, so that after-tax rates of return were virtually identical to pretax rates of return, and the $\bar{r}-g$ gap was as large as the $r-g$ gap (Figure 15.28).

The very large $\bar{r}-g$ gap until the late nineteenth to early twentieth century is in our view the primary candidate explanation as to why the concentration of wealth has been so large during most of human history. Although the rise of growth rates from less than 0.5% per year before the eighteenth century to about 1–1.5% per year during the eighteenth to

[44] For instance, simulation results suggest that differences in top inheritance tax rates can potentially explain a large fraction of the gap in wealth concentration between countries such as Germany and France (see Dell, 2005).

[45] In traditional agrarian societies, e.g., in eighteenth-century Britain or France, the market value of agricultural land was typically around 20–25 years of annual land rent, which corresponds to a rate of return of about 4–5%. Returns on more risky assets such as financial loans were sometime much higher. See Piketty (2014).

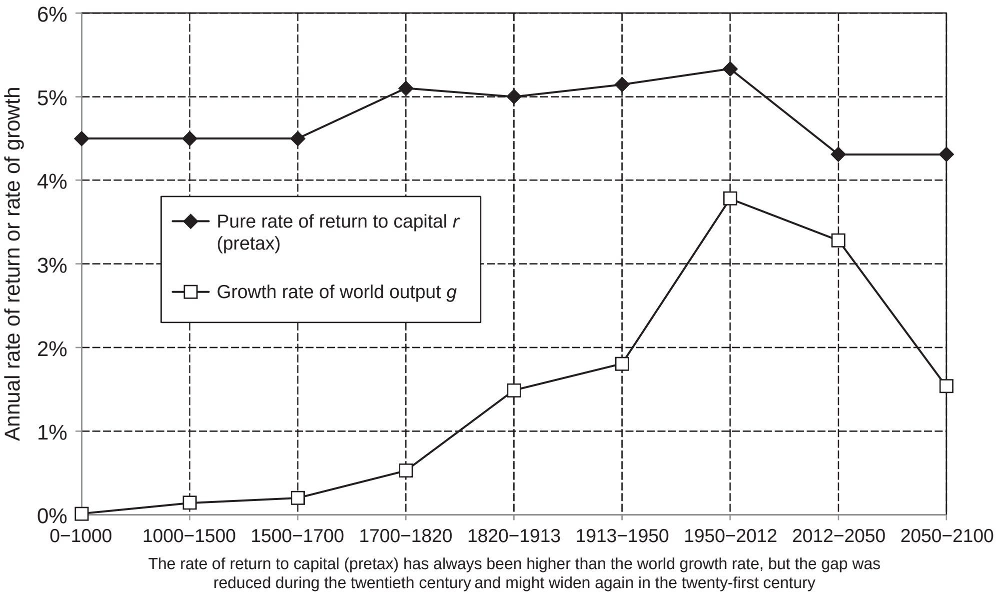


Figure 15.27 Rate of return versus growth rate at the world level, from antiquity until 2100.

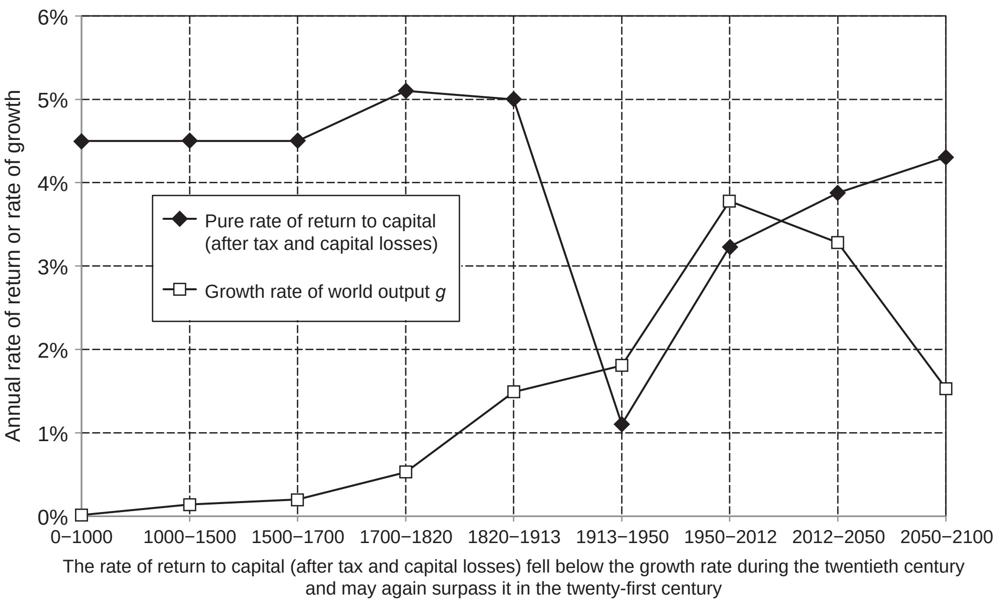


Figure 15.28 After tax rate of return versus growth rate at the world level, from antiquity until 2100.

nineteenth centuries was sufficient to make a huge difference in terms of population and living standards, it had a relatively limited impact on the $\bar{r}-g$ gap: $\bar{r}$ remained much bigger than g.[46]

The spectacular fall of the $\bar{r}-g$ gap in the course of the twentieth century can also help understand the structural decline of wealth concentration, and in particular why wealth concentration did not return to the extreme levels observed before the World Wars. The fall of the $\bar{r}-g$ gap during the twentieth century has two components: a large rise in g and a large decline in $\bar{r}$. Both, however, might well turn out to be temporary.

Start with the rise in g. The world GDP growth rate was almost 4% during the second half of the twentieth century. This is due partly to a general catch up process in per capita GDP levels (first in Europe and Japan between 1950 and 1980, and then in China and other emerging countries starting around 1980–1990), and partly to unprecedented population growth rates (which account for about half of world GDP growth rates over the past century). According to UN demographic projections, world population growth rates should sharply decline and converge to 0% during the second half of the twenty-first century. Long run per capita growth rates are notoriously difficult to predict: they might be around 1.5% per year (as posited on Figure 15.27 for the second half of the twenty-first century), but some authors—such as Gordon (2012)—believe that they could be less than 1%. In any case, it seems plausible that the exceptional growth rates of the twentieth century will not happen again—at least regarding the demographic component—and that g will indeed gradually decline during the twenty-first century.

Looking now at $\bar{r}$, we also see a spectacular decline during the twentieth century. If we take into account both the capital losses (fall in relative asset prices and physical destructions) and the rise in taxation, the net-of-tax, net-of-capital-losses rate of return $\bar{r}$ fell below the growth rate during the entire twentieth century after World War I.

Other forms of capital shocks could occur in the twenty-first century. But assuming no new shock occurs, and assuming that rising international tax competition to attract capital leads all forms of capital taxes to disappear in the course of the twenty-first century (arguably a plausible scenario, although obviously not the only possible one), the net-of-tax rate of return $\bar{r}$ will converge toward the pretax rate of return r, so that the $\bar{r}-g$ gap will again be very large in the future. Other things equal, this force could lead to rising wealth concentration during the twenty-first century.

The $\bar{r}-g$ gap was significantly larger in Europe than in the United States during the nineteenth century (due in particular to higher population growth in the New World). This fact can contribute to explain why wealth concentration was also higher in Europe. The $\bar{r}-g$ gap dramatically declined in Europe during the twentieth

[46] It is also possible that the rise of the return to capital during the eighteenth to nineteenth centuries was somewhat larger than the lower-bound estimates that we report on Figure 15.27, so that the $r-g$ gap perhaps did not decline at all. See Piketty (2014) for a more elaborate discussion.

century—substantially more than the United States, which can, in turn, explain why wealth has become structurally less concentrated than in the United States. The higher level of labor income inequality in the United States in recent decades, as well as the sharp drop in tax progressivity, also contribute to higher wealth concentration in the United States (see Saez and Zucman, 2014). Note, however, that the United States is still characterized by higher population growth (as compared to Europe and Japan), and that this tends to push in the opposite direction (i.e., less wealth concentration). So whether the wealth inequality gap with Europe will keep widening in coming decades is very much an open issue at this stage.

More generally, we should stress that although the general historical pattern of $\bar{r} - g$ (both over time and across countries) seems consistent with the evolution of wealth concentration, other factors do also certainly play an important role in wealth inequality.

One such factor is the magnitude of idiosyncratic shocks to rates of return r_{ti}, and the possibility that average rates of return $r(w) = E(r_{ti}|w_{ti} = w)$ vary with the initial wealth levels. Existing evidence on returns to university endowments suggests that larger endowments indeed tend to get substantially larger rates of returns, possibly due to scale economies in portfolio management (Piketty 2014, Chapter 12). The same pattern is found for the universe of U.S. foundations (Saez and Zucman, 2014). Evidence from *Forbes* global wealth rankings also suggests that higher wealth holders tend to get higher returns. Over the 1987–2013 period, the top fractiles (defined in proportion to world adult population) of the *Forbes* global billionaire list have been growing on average at about 6–7% per year in real terms, when average adult wealth at the global level was rising at slightly more than 2% per year (see Table 15.1).

Whatever the exact mechanism might be, this seems to indicate that the world distribution of wealth is becoming increasingly concentrated, at least at the top of the distribution. It should be stressed again, however, that available data is of relatively low quality. Little is known about how the global wealth rankings published by magazines

Table 15.1 The growth rate of top global wealth, 1987–2013

Average real growth rate per year (after deduction of inflation)	**1987–2013**
The top 1/(100 million) highest wealth holders (about 30 adults out of 3 billions in 1980s, and 45 adults out of 4.5 billions in 2010s)	6.8%
The top 1/(20 million) highest wealth holders (about 150 adults out of 3 billions in 1980s, and 225 adults out of 4.5 billions in 2010s)	6.4%
Average world wealth per adult	2.1%
Average world income per adult	1.4%
World adult population	1.9%
World GDP	3.3%

Between 1987 and 2013, the highest global wealth fractiles have grown at 6–7% per year, versus 2.1% for average world wealth and 1.4% for average world income. All growth rates are net of inflation (2.3% per year between 1987 and 2013).

are constructed, and it is likely that they suffer from various biases. They also focus on such a narrow fraction of the population that they are of limited utility for a comprehensive study of the global distribution of wealth. For instance, what happens above \$1 billion does not necessarily tell us much about what happens between \$10 and 100 million. This is a research area where a lot of progress needs to be made.

15.5.5 The Steady-State Level of the Inheritance Share: $\varphi = \varphi(g)$

15.5.5.1 The Impact of Saving Motives, Growth, and Life Expectancy

The return of high wealth–income ratios β does not necessarily imply the return of inheritance. From a purely logical standpoint, it is perfectly possible that the steady-state $\beta = s/g$ rises (say, because g goes down and s remains relatively high, as we have observed in Europe and Japan over the recent decades), but that all saving flows come from life-cycle wealth accumulation and pension funds, so that the inheritance share φ is equal to zero. Empirically, however, this does not seem to be the case. From the (imperfect) data that we have, it seems that the rise in the aggregate wealth–income ratio β has been accompanied by a rise in the inheritance share φ, at least in Europe.

This suggests that the taste for leaving bequests (and/or the other reasons for dying with positive wealth, such as precautionary motives and imperfect annuity markets) did not decline over time. Empirical evidence shows that the distribution of saving motives varies a lot across individuals. It could also be that the distribution of saving motives is partly determined by the inequality of wealth. Bequests might partly be a luxury good, in the sense that individuals with higher relative wealth also have higher bequest taste on average. Conversely, the magnitude of bequest motives has an impact on the steady-state level of wealth inequality. Take, for instance, the dynamic wealth accumulation model described above. In that model we implicitly assume that individuals leave wealth to the next generation. If they did not, the dynamic cumulative process would start at zero all over again at each generation, so that steady-state wealth inequality would tend to be smaller.

Now, assume that we take as given the distribution of bequest motives and saving parameters. Are there reasons to believe that changes in the long-run growth rate g or in the demographic parameters (such as life expectancy) can have an impact on the inheritance share φ in total wealth accumulation?

This question has been addressed by a number of authors, such as Laitner (2001) and DeLong (2003).[47] According to DeLong, the share of inheritance in total wealth accumulation should be higher in low-growth societies, because the annual volume of new savings is relatively small in such economics (so that in effect most wealth originates from inheritance). Using our notations, the inheritance share $\varphi = \varphi(g)$ is a decreasing function of the growth rate g.

[47] See also Davies et al. (2012, pp. 123–124).

This intuition is interesting (and partly correct) but incomplete. In low-growth societies, such as preindustrial societies, the annual volume of new savings—for a given aggregate β—is indeed low in steady state: $s = g \cdot \beta$. In contrast, the flow of inheritances is given by: $b_y = \mu \cdot m \cdot \beta$ (see Section 15.4). Therefore, for given μ and m, inheritance flows tend to dominate saving flows in low-growth economies, and conversely in high-growth economies.

For instance, if $\mu = 1$, $m = 2\%$, and $\beta = 600\%$, the inheritance flow is equal to $b_y = 12\%$. The inheritance flow b_y is four times bigger than the saving flow $s = 3\%$ if $g = 0.5\%$, it is equal to the saving flow $s = 12\%$ if $g = 2\%$, and it is 2.5 times smaller than the saving flow $s = 30\%$ if $g = 5\%$. Therefore—the argument goes—inherited wealth represents the bulk of aggregate wealth in low-growth, preindustrial societies; makes about half of aggregate wealth in medium-growth, mature economies; and a small fraction of aggregate wealth in high-growth, booming economies.

This intuition, however, is incomplete, for two reasons. First, as we already pointed out in Section 15.4, saving flows partly come from the return to inherited wealth, and this needs to be taken into account. Next, the μ parameter, i.e., the relative wealth of decedents, is endogenous and might well depend on the growth rate g, as well as on demographic parameters such as life expectancy and the mortality rate m. In the pure life-cycle model where agents die with zero wealth, μ is always equal to zero, and so is the inheritance share φ, independently of the growth rate g, no matter how small g is. For given (positive) bequest tastes and saving parameters, however, one can show that in steady state, $\mu = \mu(g, m)$ tends to be higher when growth rates g and mortality rates m are lower.

15.5.5.2 A Simple Benchmark Model of Aging Wealth and Endogenous μ

To see this point more clearly, it is necessary to put more demographic structure into the analysis. Here we follow a simplified version of the framework introduced by Piketty (2011).

Consider a continuous-time, overlapping-generations model with a stationary population $N_t = [0, 1]$ (zero population growth). Each individual i becomes adult at age $a = A$, has exactly one child at age $a = H$, and dies at age $a = D$. We assume away inter-vivos gifts, so that each individual inherits wealth solely when his or her parent dies, that is, at age $a = I = D - H$.

For example, if $A = 20$, $H = 30$, and $D = 60$, everybody inherits at age $I = D - H = 30$ years old. But if $D = 80$, then everybody inherits at age $I = D - H = 50$ years old.

Given that population N_t is assumed to be stationary, the (adult) mortality rate m_t is also stationary, and is simply equal to the inverse of (adult) life expectancy: $m_t = m = \frac{1}{D-A}$.[48]

[48] It is more natural to focus upon adults because minors usually have very little income or wealth (assuming that $I > A$, i.e., $D - A > H$, which is the case in modern societies).

For example, if $A=20$ and $D=60$, the mortality rate is $m=1/40=2.5\%$. If $D=80$, the mortality rate is $m=1/60=1.7\%$. That is, in a society where life expectancy rises from 60 to 80 years old, the steady-state mortality rate among adults is reduced by a third. In the extreme case where life expectancy rises indefinitely, the steady-state mortality rate becomes increasingly small: one almost never dies.

Does this imply that the inheritance flow $b_y=\mu\cdot m\cdot\beta$ will become increasingly small in aging societies? Not necessarily: even in aging societies, one ultimately dies. Most importantly, one tends to die with higher and higher relative wealth. That is, wealth also tends to get older in aging societies, so that the decline in the mortality rate m can be compensated by a rise in relative decedent wealth μ (which, as we have seen, has been the case in France).

Assume for simplicity that all agents have on average the same uniform saving rate s on all their incomes throughout their life (reflecting their taste for bequests and other saving motives such a precautionary wealth accumulation) and a flat age-income profile (including pay-as-you-go pensions). Then one can show that the steady-state $\mu=\mu(g)$ ratio is given by the following formula:

$$\mu(g)=\frac{1-\mathrm{e}^{-(g-s\cdot r)(D-A)}}{1-\mathrm{e}^{-(g-s\cdot r)H}}=\frac{1-\mathrm{e}^{-(1-\alpha)g(D-A)}}{1-\mathrm{e}^{-(1-\alpha)gH}}.$$

With $\alpha=r\cdot\beta=r\cdot s/g=$ capital share in national income.

In other words, the relative wealth of decedents $\mu(g)$ is a decreasing function of the growth rate g (and an increasing function of the rate of return r or of the capital share α).[49] If one introduces taxes into the model, one can easily show that μ is a decreasing function of the growth rate g and an increasing function of net-of-tax rate of return $\bar{r}$ (or the net-of-tax capital share $\overline{\alpha}$).[50]

The intuition for this formula, which can be extended to more general saving models, is the following. With high growth rates, today's incomes are large as compared to past incomes, so the young generations are able to accumulate almost as much wealth as the older cohorts, in spite of the fact that the latter have already started to accumulate in the past, and in some cases have already received their bequests. Generally speaking, high growth rates g are favorable to the young generations (who are just starting to accumulate wealth, and who therefore rely entirely on the new saving flows out of current incomes), and tend to push for lower relative decedent wealth μ. High rates of return $\bar{r}$, by contrast,

[49] This steady-state formula applies both to the closed-economy and open-economy cases. The only difference is that the rate of return r is endogenously determined by the marginal product of domestic capital accumulation in the closed economy case (e.g., $r=F_K=a\cdot\beta^{-1/\sigma}$ with a CES production function), while it is a free parameter in the open economy setup (in which case the formula can be viewed as $\mu=\mu(g, r)$).

[50] With taxes, $\bar{r}$ also becomes a free parameter in the closed-economy model, so the formula should always be viewed as $\mu=\mu(g,\bar{r})$.

are more favorable to older cohorts, because this makes the wealth holdings that they have accumulated or inherited in the past grow faster, and tend to pusher for higher μ.

In the extreme case where $g \to \infty$, then $\mu \to 1$ (this directly follows from flat saving rates and age-labor income profiles).

Conversely, in the other extreme case where $g \to 0$, then $\mu \to \overline{\mu} = \frac{D-A}{H} > 1$.

It is worth noting that this maximal value $\overline{\mu}$ rises in proportion to life expectancy $D-A$ (for given generation length H). Intuitively, with $g \approx 0$ and uniform saving, most of wealth originates from inheritance, so that young agents are relatively poor until inheritance age $I = D-H$, and most of the wealth concentrates between age $D-H$ and D, so that relative decedent wealth $\mu \approx \overline{\mu} = \frac{D-A}{H}$.[51]

That is, as life expectancy $D-A$ rises, wealth gets more and more concentrated at high ages. This is true for any growth rate, and all the more for low growth rates. In aging societies, one inherits later in life,[52] but one inherits bigger amounts. With $g \approx 0$, one can see that both effects exactly compensate each other, in the sense that the steady-state inheritance flow b_y is entirely independent of life expectancy. That is, with $m = \frac{1}{D-A}$ and $\overline{\mu} = \frac{D-A}{H}$, we have $b_y = \overline{\mu} \cdot m \cdot \beta = \frac{\beta}{H}$, independently from $D-A$. For a given wealth–income ratio $\beta = 600\%$ and generation length $H = 30$ years, the steady-state annual inheritance flow is equal to $b_y = 20\%$ of national income, whether life expectancy is equal to $D = 60$ years or $D = 80$ years.

Strictly speaking, this is true only for infinitely small growth $g \approx 0$. However, by using the above formula one can see that for low growth rates (say, $g \approx 1$–1.5%) then the steady-state inheritance flow is relatively close to $b_y = \frac{\beta}{H}$ and is almost independent of life expectancy. It is only for high growth rates—above 2–3% per year—that the steady-state inheritance flow is reduced substantially.

15.5.5.3 Simulating the Benchmark Model

Available historical evidence shows that the slowdown of growth is the central economic mechanism explaining why the inheritance flow seems to be returning in the early twenty-first century to approximately the same level $b_y \approx 20\%$ as that observed during the nineteenth and early twentieth centuries.

By simulating a simple uniform-saving model for the French economy over the 1820–2010 period (starting from the observed age–wealth pattern in 1820, and using observed aggregate saving rates, growth rates, mortality rates, capital shocks and age–labor

[51] In the extreme case where young agents have zero wealth and agents above age $I = D-H$ have average wealth $\overline{w}$, then average wealth among the living is equal to $w = \frac{(D-I) \cdot \overline{w}}{D-A}$ and, so that $\overline{\mu} = \frac{\overline{w}}{w} = \frac{D-A}{H}$. See Piketty (2011), Propositions 1–3.

[52] Although in practice, this is partly undone by the rise of inter vivos gifts, as we have seen above.

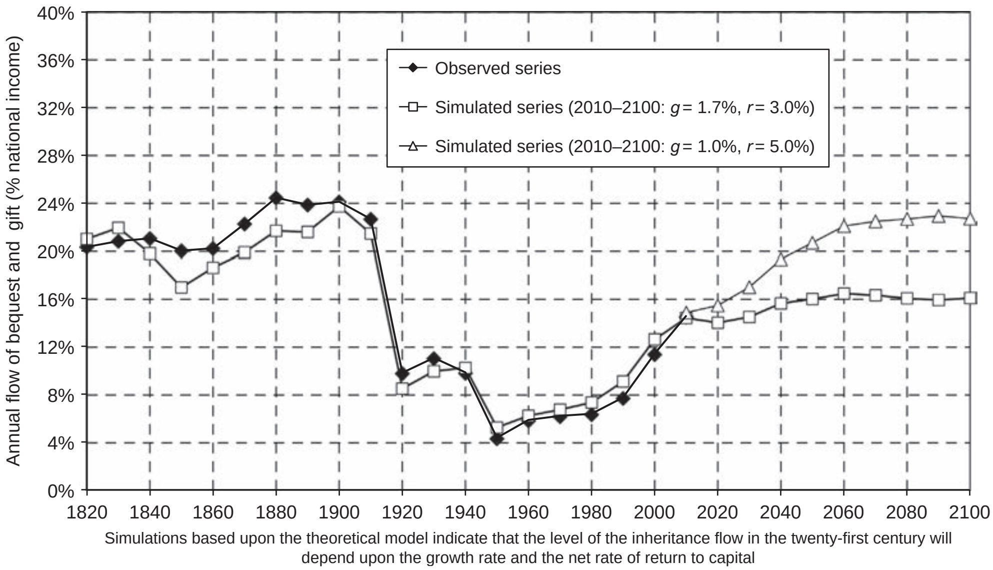


Figure 15.29 Observed and simulated inheritance flow, France 1820–2100.

income profiles over the entire period), one can reasonably well reproduce the dynamics of the age–wealth profile and hence of the μ ratio and the inheritance flow b_y over almost two centuries (see Figure 15.29).

We can then use this same model to simulate the future evolution of the inheritance flow in coming decades. As one can see on Figure 15.29, a lot depends on future values of the growth rate g and the net-of-tax rate of return $\bar{r}$ over the 2010–2100 period. Assuming $g = 1.7\%$ (which corresponds to the average growth rate observed in France between 1980 and 2010) and $\bar{r} = 3.0\%$ (which approximatively corresponds to net-of-tax average real rate of return observed in 2010), then b_y should stabilize around 16–17% in coming decades. If growth slows $g = 1.0\%$ and the net-of-tax rate of return rises to $\bar{r} = 5.0\%$ (e.g., because of a rise of the global capital share and rate of return, or because of a gradual repeal of capital taxes), b_y would keep increasing toward 22–23% over the course of the twenty-first century. The flow of inheritance would approximately return to its nineteenth and early twentieth centuries level.

In Figure 15.30, we use these projections to compute the corresponding share φ of cumulated inheritance in the aggregate wealth stock (using the PPVR definition and the same extrapolations as those described above). In the first scenario, φ stabilizes around 80%; in the second scenario, it stabilizes around 90% of aggregate wealth.

These simulations, however, are not fully satisfactory, first because a lot more data should be collected on inheritance flows in other countries, and next because one should ideally try to analyze and simulate both the flow of inheritance and the inequality of

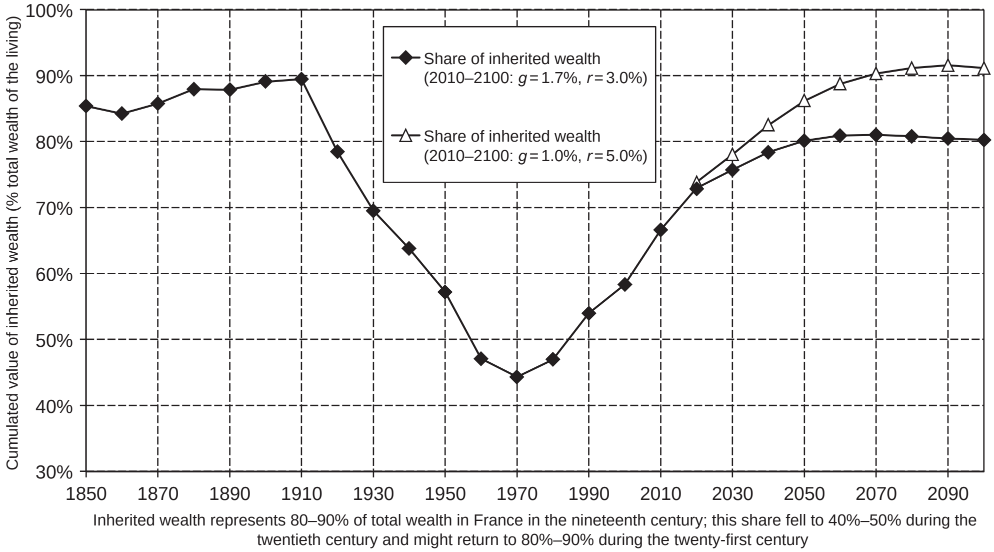


Figure 15.30 The share of inherited wealth in total wealth, France 1850–2100.

wealth. The computations presented here assume uniform saving and solely attempt to reproduce the age-average wealth profile, without taking into account within-cohort wealth inequality. This is a major limitation.

15.6. CONCLUDING COMMENTS AND RESEARCH PROSPECTS

In this chapter, we have surveyed the empirical and theoretical literature on the long-run evolution of wealth and inheritance in relation to output and income. The magnitude and concentration of wealth and inheritance (relative to national income) were very high in the eighteenth to nineteenth centuries up until World War I, then dropped precipitously during the twentieth century following World War shocks, and have been rising again in the late twentieth and early twenty-first centuries. We have showed that over a wide range of models, the long-run magnitude and concentration of wealth and inheritance are an increasing function of $\bar{r}-g$, where $\bar{r}$ is the net-of-tax rate of return to wealth and g is the economy's growth rate, and we have argued that these predictions are broadly consistent with historical patterns. These findings suggest that current trends toward rising wealth–income ratios and wealth inequality might continue during the twenty-first century, both because of the slowdown of population and productivity growth, and because of increasing international competition to attract capital.

We should stress, however, that this is an area where a lot of progress still needs to be made. Future research should particularly focus on the following issues. First, it becomes more and more important to study the dynamics of the wealth distribution from a global perspective.[53] In order to do so, it is critical to take into account existing macro data on aggregate wealth and foreign wealth holdings. Given the large movements in aggregate wealth–income ratios across countries, such macro-level variations are likely to have a strong impact on the global dynamics of the individual-level distribution of wealth. It is also critical to use existing estimates of offshore wealth and to analyze how much tax havens are likely to affect global distributional trends (see Zucman, 2014). Next, a lot more historical and international data needs to be collected on inheritance flows. Last, there is a strong need of a better articulation between empirical and theoretical research. A lot more work has yet to be done before we are able to develop rigorous and credible calibrations of dynamic theoretical models of wealth accumulation and distribution.

ACKNOWLEDGMENTS

We are grateful to the editors and to Daniel Waldenstrom for helpful comments. All series and figures presented in this chapter are available in an online data appendix.

REFERENCES

Antras, P., 2004. Is the U.S. aggregate production function Cobb-Douglas? New estimates of the elasticity of substitution. B.E. J. Macroecon. 4 (1) article 4.
Atkinson, A., 1983. The Economics of Inequality. Clarendon Press, Oxford, 330 pp.
Atkinson, A.B., 2013. Wealth and Inheritance in Britain from 1896 to the Present. Working Paper.
Atkinson, A., Harrison, A.J., 1978. Distribution of Personal Wealth in Britain, 1923–1972. Cambridge University Press, Cambridge, 330 pp.
Atkinson, A.B., Piketty, T., Saez, E., 2011. Top incomes in the long run of history. J. Econ. Lit. 49 (1), 3–71.
Benhabib, J., Bisin, A., Zhu, S., 2011. The distribution of wealth and fiscal policy in economies with finitely lived agents. Econometrica 79 (1), 123–157.
Benhabib, J., Bisin, A., Zhu, S., 2013. The Wealth Distribution in Bewley Models with Investment Risk. NYU, New York.
Blinder, A., 1988. Comments on Modigliani and Kotlikoff-Summers. In: Kessler, D., Masson, A. (Eds.), Modelling the Accumulation and Distribution of Wealth. Oxford University Press, Oxford, pp. 68–76.
Boisguillebert, P., 1695. Le détail de la France, Paris, 215 pp. Reprinted in E. Daire, Economistes financiers du 18e siècle, Paris: Guillaumin, 1843.
Bowley, A.L., 1920. The Change in the Distribution of National Income, 1880–1913. Clarendon Press, Oxford.
Champernowne, D.G., 1953. A model of income distribution. Econ. J. 63 (250), 318–351.
Chirinko, R., 2008. σ: the long and short of it. J. Macroecon. 30, 671–686.
Colquhoun, P., 1815. A Treatise on the Wealth. Power and Resources of the British Empire, London, 561 pp.
Colson, C., 1903. Cours d'économie politique. Gauthier-Villars, Paris, 1918, 1927 (several editions).
Cowell, F., 1998. Inheritance and the Distribution of Wealth. LSE, London.
Davies, J.B., Shorrocks, A.F., 1999. The distribution of wealth. In: Atkinson, A.B., Bourguignon, F. (Eds.), Handbook of Income Distribution, vol. 1. pp. 605–675 (Chapter 11).

[53] See the important pioneering work of Davies et al. (2010, 2012).

Davies, J.B., Sandström, S., Shorrocks, A., Wolff, E.N., 2010. The level and distribution of global household wealth. Econ. J. 121, 223–254.
Davies, J.B., Lluberas, R., Shorrocks, A., 2012. Global Wealth Data-book. Credit Suisse. https://www.credit-suisse.com/ch/en/news-and-expertise/research/credit-suisse-research-institute/publications.html.
Dell, F., 2005. Top incomes in Germany and Switzerland over the twentieth century. J. Eur. Econ. Assoc. 3, 412–421.
de Foville, A., 1893. The wealth of France and of other countries. J. R. Stat. Soc. 56 (4), 597–626.
DeLong, J.B., 2003. Bequests: an historical perspective. In: Munnel, A. (Ed.), The Role and Impact of Gifts and Estates. Brookings Institution, Washington, DC.
Edlund, L., Kopczuk, W., 2009. Women, wealth and mobility. Am. Econ. Rev. 99 (1), 146–178.
Fisher, I., 1920. Economists in public service. Am. Econ. Rev. 9, 5–21.
Giffen, R., 1889. The Growth of Capital. G. Bell and sons, London, 169 pp.
Goldsmith, R.W., 1985. Comparative National Balance Sheets: A Study of Twenty Countries, 1688–1978. The University of Chicago Press, Chicago, 353 pp.
Goldsmith, R.W., 1991. Pre-Modern Financial Systems. Cambridge University Press, Cambridge, 348 pp.
Gordon, R.J., 2012. Is U.S. Economic Growth Over? Faltering Innovation Confronts the Six Headwinds. NBER Working Paper 18315.
Greenwood, D., 1983. An estimation of U.S. family wealth and its distribution from microdata, 1973. Rev. Income Wealth 29, 23–43.
Helfferich, K., 1913. Deutschlands Volkswholhstand, 1888–1913. G. Stilke, Berlin.
Hopehnayn, H., Prescott, E., 1992. Stochastic monotonicity and stationary distributions for dynamic economies. Econometrica 60 (6), 1387–1406.
Jones, C.I., Romer, P.M., 2010. The new Kaldor facts: ideas, institutions, population, and human capital. Am. Econ. J. 2 (1), 224–245.
Kaldor, N., 1961. Capital accumulation and economic growth. In: Lutz, F.A., Hague, D.C. (Eds.), The Theory of Capital. Saint Martins Press, New York.
Karabarbounis, L., Neiman, B., 2014. The global decline of the labor share. Q. J. Econ. 129 (1), 61–103.
King, G., 1696. Natural and Political Observations and Conclusions Upon the State and Condition of England, 1696. pp. 29–73, 45 pp., Published in Chalmers 1804.
King, W.I., 1915. The Wealth and Income of the People of the United States. McMillan Press, New York , 278 pp.
King, W.I., 1927. Wealth distribution in the Continental United States at the close of 1921. J. Am. Stat. Assoc. 22, 135–153.
Kaplan, S.N., Rauh, J., 2013. Family, education, and sources of wealth among the richest Americans, 1982–2012. Am. Econ. Rev. 103 (3), 158–166.
Karagiannaki, E., 2011. Recent Trends in the Size and the Distribution of Inherited Wealth in the UK. London School of Economics, London, CASE Paper 146.
Kopczuk, W., Saez, E., 2004. Top wealth shares in the United States, 1916–2000: evidence from estate tax returns. Natl. Tax J. 57 (2), 445–487.
Kotlikoff, L., 1988. Intergenerational transfers and savings. J. Econ. Perspect. 2 (2), 41–58.
Kotlikoff, L., Summers, L., 1981. The role of intergenerational transfers in aggregate capital accumulation. J. Polit. Econ. 89, 706–732.
Kuznets, S., 1953. Shares of upper income groups in income and savings, 1913–1948. National Bureau of Economic Research, Cambridge, MA, 707 pp.
Lampman, R.J., 1962. The share of top wealth-holders in national wealth 1922–1956. Princeton University Press, Princeton.
Laitner, J., 1979. Bequest behaviour and the national distribution of wealth. Rev. Econ. Stud. 46 (3), 467–483.
Laitner, J., 2001. Secular changes in wealth inequality and inheritance. Econ. J. 111 (474), 691–721.
Lindert, P., 1986. Unequal English wealth since 1688. J. Polit. Econ. 94 (6), 1127–1162.
Lindert, P., 2000. Three centuries of inequality in Britain and America. In: Atkinson, A.B., Bourguignon, F. (Eds.), Handbook of Income Distribution, vol. 1. Elsevier/North Holland, Amsterdam, pp. 167–216.

Maddison, A., 2010. Historical Statistics of the World Economy: 1–2008 AD. www.ggcc.net/maddisonupdated tables.
Mallet, B., 1908. A method of estimating capital wealth from the estate duty statistics. J. R. Stat. Soc. 71 (1), 65–101.
Mallet, B., Strutt, H.C., 1915. The multiplier and capital wealth. J. R. Stat. Soc. 78 (4), 555–599.
Meade, J.E., 1964. Efficiency, Equality and the Ownership of Property. Allen and Unwin, London.
Menchik, P.L., 1980. Primogeniture, equal sharing, and the U.S. distribution of wealth. Q. J. Econ. 94 (2), 299–316.
Modigliani, F., 1986. Life cycle, individual thrift and the wealth of nations. Am. Econ. Rev. 76 (3), 297–313.
Modigliani, F., 1988. The role of intergenerational transfers and lifecycle savings in the accumulation of wealth. J. Econ. Perspect. 2 (2), 15–40.
Morigushi, C., Saez, E., 2008. The evolution of income concentration in Japan, 1886–2005: evidence from income tax statistics. Rev. Econ. Stat. 90 (4), 713–734.
Nirei, M., 2009. Pareto Distributions in Economic Growth Models. Hitotsubashi University, Tokyo, Working Paper.
Ohlsson, H., Roine, J., Waldenstrom, D., 2013. Inherited Wealth over the Path of Development: Sweden 1810-2010. Mimeo, Uppsala.
Petty, W., 1664. Verbum Sapienti. 26 pp. Published as an addendum to Petty, W., 1691. The Political Anatomy of Ireland. London, 266 pp. Reprinted in The Economic Writings of Sir William Petty, Hull, C.H. (Ed.), 2 vols., Cambridge University Press, 1989.
Piketty, T., 2011. On the long-run evolution of inheritance: France 1820–2050. Q. J. Econ. 126 (3), 1071–1131.
Piketty, T., 2014. Capital in the 21st Century. Harvard University Press, Cambridge.
Piketty, T., Saez, E., 2013. A theory of optimal inheritance taxation. Econometrica 81 (5), 1851–1886.
Piketty, T., Zucman, G., 2014. Capital is back: wealth-income ratios in rich countries 1700–2010. Q. J. Econ. 129 (3), 1255–1310.
Piketty, T., Postel-Vinay, G., Rosentha, J.-L., 2006. Wealth concentration in a developing economy: Paris and France, 1807–1994. Am. Econ. Rev. 96 (1), 236–256.
Piketty, T., Postel-Vinay, G., Rosenthal, J.-L., 2013. Inherited versus self-made wealth: theory and evidence from a rentier society (1872–1927). Explor. Econ. Hist. 51, 21–40.
Roine, J., Waldenstrom, D., 2009. Wealth concentration over the path of development: Sweden, 1873–2006. Scand. J. Econ. 111, 151–187.
Saez, E., Zucman, G., 2014. Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data. Working Paper.
Schinke, C., 2013. Inheritance in Germany 1911 to 2009. PSE Working Paper.
Séaillès, J., 1910. La répartition des fortunes en France. Editions Felix Alcan, Paris, 143 pp.
Stamp, J.C., 1919. The wealth and income of the chief powers. J. R. Stat. Soc. 82 (4), 441–507.
Stewart, C., 1939. Income capitalization as a method of estimating the distribution of wealth by size group. Studies in Income and Wealth, vol. 3. National Bureau of Economic Research, New York.
Stiglitz, J.E., 1969. Distribution of income and wealth among individuals. Econometrica 37 (3), 382–397.
Strutt, H.C., 1910. Notes on the distribution of estates in France and the United Kingdom. J. R. Stat. Soc. 73 (6), 634–644.
Vauban, S., 1707. Projet d'une dixme royale. Librairie Félix Alcan, Paris, 267 pp. Reprinted in E. Daire, Economistes financiers du 18e siècle, Paris: Guillaumin, 1843.
Vaughan, R.N., 1979. Class behaviour and the distribution of wealth. Rev. Econ. Stud. 46 (3), 447–465.
Waldenstrom, D., 2009. Lifting All Boats? The Evolution of Income and Wealth Inequality over the Path of Development. PhD Dissertation, Lunds universitet, Lund.
Wolff, E., Gittleman, M., 2013. Inheritances and the distribution of wealth or whatever happened to the great inheritance boom? J. Econ. Inequal. 1–30, available online at http://dx.doi.org/10.1007/s10888-013-9261-8.
Zucman, G., 2013. The missing wealth of nations: are Europe and the U.S. net debtors or net creditors? Q. J. Econ. 128 (3), 1321–1364.
Zucman, G., 2014. Taxing Across Borders: Tracking Personal Wealth and Corporate Profits. J. Econ. Perspect. forthcoming.

CHAPTER 16

Intrahousehold Inequality

Pierre-André Chiappori*, Costas Meghir†
*Department of Economics, Columbia University, NY, USA
†Department of Economics, Yale University, New Haven, CT, USA

Contents

Abstract

Studies of inequality often ignore resource allocation within the household. In doing so they miss an important element of the distribution of welfare that can vary dramatically depending on overall environmental and economic factors. Thus, measures of inequality that ignore intrahousehold allocations are both incomplete and misleading. We discuss determinants of intrahousehold allocation of

Handbook of Income Distribution, Volume 2B
ISSN 1574-0056, http://dx.doi.org/10.1016/B978-0-444-59429-7.00017-0

resources and welfare. We show how the sharing rule, which characterizes the within-household allocations, can be identified from data on household consumption and labor supply. We also argue that a measure based on estimates of the sharing rule is inadequate as an approach that seeks to understand how welfare is distributed in the population because it ignores public goods and the allocation of time to market work, leisure, and household production. We discuss a money metric alternative, that fully characterizes the utility level reached by the agent. We then review the current literature on the estimation of the sharing rule based on a number of approaches, including the use of distribution factors as well as preference restrictions.

Keywords

Collective model, Income distribution, Economics of the family, Labor supply, Household behavior

JEL Classification Codes

D1, D11, D12, D13, D31, D6, H41

16.1. INTRODUCTION

16.1.1 Inequality Between Individuals

Consider an economy with identical couples, each of whom has a total income of 100. Individuals privately consume a unique, perfectly divisible commodity; there exist neither externalities nor economies of scale, so that, in each couple, the sum of individual consumptions equals the couple's total income. Inequality, as measured in a standard way, is nil. Assume, now, that some of these couples divorce, and that after divorce husbands each receive an income of 75, while each wife gets 25. The new income distribution, again by standard criteria, is now unequal; in particular, the presence of lower-income singles (the divorced wives) increases both inequality and poverty.

From a deeper perspective, however, the conclusion just stated is far from granted. It entirely relies on an implicit assumption—namely, that the predivorce distribution of income *within households* was equal. Most of the time, such an assumption has little or no empirical justification; and from a theoretical viewpoint, it is actually quite unlikely to hold—few serious models of household behavior would predict an equal distribution of income while married if the post-divorce allocation is highly skewed. Still, it is crucial. Assume, for the sake of the argument, that the distribution of resources within married couples simply mimics what it would be in case of divorce (he gets 75, she gets 25)—not an unreasonable assumption, given that in our (admittedly simplistic) structure this is the only individually rational allocation. Then the claim that inequality increased after the wave of divorces is simply wrong. Inequality, at least across individuals, has not changed; each agent has exactly the same income, consumption and welfare than before. And the surge in measured poverty is just as spurious. There are exactly as many poor women after than there were before; it is just that, in the predivorce situation, the standard measures missed them.

The previous example, extreme as it may be, illustrates a basic point that this chapter will try to emphasize—namely, that any attempt at measuring inequality (or its evolution

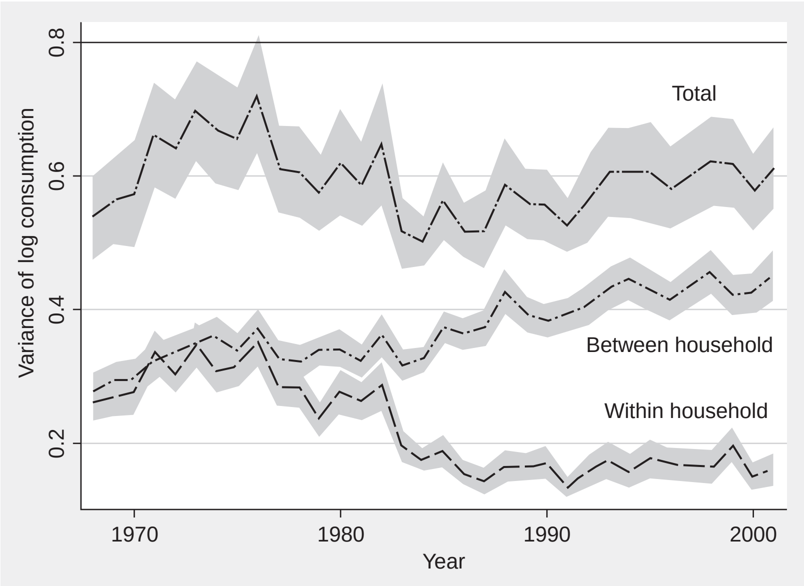


Figure 16.1 Trends in the variance of log consumption, UK. *Source:* Lise and Seitz (2011).

over time) that ignores allocation of resources within the family is unreliable at best, and deeply flawed at worst, especially when the basic demographics regarding family composition evolve over the period under consideration. This point had already been emphasized in the literature; for instance, Haddad and Kanbur (1990) have showed, on Philippine data, that standard measures of inequality in calorie adequacy would be understated by 30–40% if intrahousehold inequality was ignored. As a more recent illustration, consider the graph in Figure 16.1, due to Lise and Seitz (2011), that plots the evolution of inequality across households, within households, and across individuals in the United Kingdom over the last decades, as estimated from a collective model of labor supply. The main conclusion is that the standard approach, based on adult equivalence scales, underestimates the initial level of cross-sectional consumption inequality by 50%. Moreover, it gives a deeply flawed picture of the evolution of inequality over the last decades. While the usual story—a large surge in inequality between 1970 and 2000—applies to inequality across households, it is compensated by a considerable reduction of intrahousehold inequality, so that total inequality (across *individuals*) remains more or less constant over the period.[1]

[1] This conclusion must be qualified in view of the population under consideration. Indeed, the sample excludes all households with children, all persons aged under 22 or over 65, all persons who were self-employed, and the top 1% of the earnings distribution (which is, in any case, not well covered by the Family Expenditure Survey); so the conclusions are only valid for that particular subpopulation. Still, it is suggestive of the general claim that ignoring intrahousehold allocation may severely bias our views regarding inequality.

All this strongly suggests, at the very least, that much more attention should be paid to intrahousehold inequality, from both a theoretical and an empirical viewpoint. Analyzing intrahousehold inequality, however, raises a host of specific problems. Some are of a conceptual nature. A large fraction of household expenditures relate to public commodities, that is, goods that are jointly consumed by the household, without exclusion restrictions; moreover, in many cases these public commodities are internally produced within the household. Spouses may have different preferences regarding public goods; therefore, the fraction of household expenditures devoted to public consumption has a potentially important impact on intrahousehold inequality, that cannot be disregarded. Similar questions arise for intrahousehold production, with the additional twist that time spent by each spouse should also be taken into account. How should such public productions and consumptions be taken into account in our inequality measures? Although the impact of public goods on inequality is by no means a new problem, it is particularly stringent in our context, if only because public goods and domestic production are among the main (economic) reasons for the existence of the household.

As we shall see, these conceptual issues affect the standard notion of inequality in two ways. Besides shedding a new light on its measurement, they also revive some old discussions about its foundations. In particular, the role of public goods raises questions about which type on inequality we should concentrate on: income? (private) consumption? utility? The problem is far from innocuous: in the presence of public goods, it is relatively easy to generate examples in which a change in prices and incomes results in a decrease in a person's private consumption and an increase in the spouse's, whereas utilities evolve in the opposite way (welfare declines for the person whose private consumption increases and conversely). In such a context, the impact of the change on intrahousehold inequality is not clearly defined: It all depends on what exactly we are interested in.

Empirical problems are equally challenging. As always in economics, preferences are not directly observed and have to be recovered from observable data (demand, labor supply). But, in addition, the allocation of resources within the household cannot (in general) be directly observed; it has to be recovered from the household's (aggregate) behavior. It follows that when deciding which aspect of inequality should be considered, one cannot abstract from identification issues: there is little interest in concentrating on a notion that is not identifiable in practice. An interesting paradox, in this respect, is provided by a standard result of household economics, namely, that in some circumstances, a continuum of different models generate the same observable behavior (so they are observationally indistinguishable). In some cases, these models correspond to different intrahousehold allocations of resources, but to the same allocation of utility (in the language of the theory, the indeterminacy is welfare-irrelevant). In other words, the main justification for concentrating on inequality in income or consumption rather than in utility—namely, the fact that the former are observed, but not the latter—is sometimes partially reversed.

These questions obviously arise whenever inequality is assessed on a utilitarian basis. However, even an alternative approach in terms of capabilities could hardly disregard them. Issues related to individual preferences for public goods would be less problematic in that case; what matters, from a capabilities perspective, is more an individual's potential access to the public goods than the utility the individual actually derives from their consumption. But the difficulties in recovering individual private consumptions (especially when it comes to nutrition or other fundamental needs) would become all the more crucial. All in all, the problems raised by intrahousehold allocation should be central to any analysis of inequality, even though specific aspects may be more damaging for some approaches than for others.

What recent developments in the literature clearly indicate, however, is that while these problems are serious, they are by no means insuperable. Although intrahousehold allocation is not (fully) observable, it can be recovered using specific, identifying assumptions that will be discussed later; that is the path followed by Lise and Seitz, but also by Chiappori et al. (2002), Dunbar et al. (2013), Browning et al. (2013), and many others in the literature. Clear progress has been made on this front over the last decades. One goal of this chapter is to briefly review these advances.

A first step is to adopt an explicit model of household decision making that clarifies the notion of inequality within the household. Obviously, such models must explicitly recognize that household members each have their own preferences—if only because omitting individuals does not seem a promising way of analyzing inequality between them. An additional requirement is empirical tractability. To be usable, a model of household behavior should fulfill a double requirement: testability (i.e., it should generate a set of empirically testable restrictions that fully characterize the model, in the sense that any given behavior is compatible with the model if and only if these conditions are satisfied) and identifiability (it should be feasible, possibly under additional assumption, to recover the structure of the model—in our case, individual preferences and the decision process—from the sole observation of household behavior). Lastly, the model should provide (or be compatible with) an "upstream" theory of the generation of intrahousehold inequality; that is, we need to explain, and ideally predict, how the intrahousehold distribution of resources—and ultimately of power—responds to changes in the household's socioeconomic environment.

Most of the recent advances use one particular class of models, based on the collective approach (see Chiappori, 1988, 1992).[2] Although other (nonunitary) perspectives have been adopted in the literature, none of the alternatives has (so far) convincingly addressed the double requirement of testability and identifiability just evoked.

[2] For a more detailed presentation, the reader is referred to Browning et al. (2014).

16.1.2 Modeling Household Decision Making: The Collective Model

The basic axiom of the collective approach is Pareto efficiency: Whatever decision the household is making, no alternative choice would have been preferred by *all* members. Whereas this assumption is undoubtedly restrictive, its scope remains quite large. It encompasses as particular cases many models that have been proposed in the literature, including:

- "Unitary" models, which posit that the household behaves like a single decision maker; this includes simple dictatorship (possibly by a "benevolent patriarch," as in Becker, 1974) to the existence of some household welfare function (as in Samuelson, 1956).
- Models based on cooperative game theory, and particularly bargaining theory (at least in a context of symmetric information), as pioneered by Manser and Brown (1980) and McElroy and Horney (1981).
- Model based on market equilibrium, as analyzed by Grossbard-Shechtman (1993), Gersbach and Haller (2001), Edlund and Korn (2002), and others.
- More specific models, such as Lundberg and Pollak's "separate spheres" (1993) framework.

On the other hand, the collective framework excludes models based on noncooperative game theory (at least in the presence of public good), such as those considered by Ulph (2006), Browning et al. (2010), Lechene and Preston (2011), and many others, as well as models of inefficient bargaining a la Basu (2006).

The efficiency assumption is standard in many economic contexts and has often been applied to household behavior. Still, it needs careful justification. Within a static context, this assumption amounts to the requirement that married partners will find a way to take advantage of opportunities that make both of them better off. Because of proximity and durability of the relation, both partners are in general aware of the preferences and actions of each other. They can act cooperatively by reaching some binding agreement. Enforcement of such agreements can be achieved through mutual care and trust, by social norms and by formal legal contracts. Alternatively, the agreement can be supported by repeated interactions, including the possibility of punishment. A large literature in game theory, based on several "folk theorems," suggests that in such situations, efficiency should prevail.[3] At the very least, efficiency can be considered as a natural benchmark.

Another potential issue with a collective approach to inequality issues is of a more conceptual nature. By definition, the collective approach is axiomatic; it assumes specific properties of the outcome (efficiency), and leaves aside the specific process by which this outcome has been generated. It has sometimes been argued that one should judge differently situations that generate the same allocations (and the same utility levels) but

[3] Note, however, that folk theorems essentially apply to infinitely repeated interactions.

which are reached by different processes. In that case, the collective approach has to be further specialized, and this may be (and has been) done in several directions.[4]

Finally, an obvious but crucial advantage of the collective model is that it has been by now fully characterized. We have a set of necessary and sufficient conditions for a demand function to stem from a collective framework (Chiappori and Ekeland 2006); exclusion restrictions have been derived under which individual preferences and the decision process (as summarized by the Pareto weights) can be recovered from the sole observation of household behavior (Chiappori and Ekeland, 2009a,b). To the best of our knowledge, this is the only model of the household for which similar results have been derived.[5]

The next section describes the basic model. We then discuss the conceptual issues linked with intrahousehold inequality, first in the case where all commodities are privately consumed, then in the presence of public goods, finally for the case of domestic production. We then discuss the determinants of intrahousehold allocations followed by a section on identification of preferences and the sharing rule. Finally, we discuss issues related to identification. In the following section we give an overview of empirical findings and then we conclude with a brief discussion of future directions of research.

16.2. THE COLLECTIVE MODEL: CONCEPTS, DEFINITIONS, AND AXIOMS

In what follows, we consider a K-person household that can consume several commodities; these may include standard consumption goods and services, but also leisure, future or contingent goods, and the like. Formally, N of these commodities are publicly consumed within the household. The market purchase of public good j is denoted Q_j; the N-vector of public goods is given by Q. Similarly, private goods are denoted q_i with the n-vector q. Each private good bought is divided between the members so that member a ($a=1,\ldots,K$) receives q_i^a of good i, with $\sum_a q_i^a = q_i$. The vector of private goods that a receives is q^a, with $\sum_a q^a = q$. An *allocation* is an $N+Kn$-vector $(Q, q^1, \ldots, q^K)$. The associated market prices are given by the N-vector P and the n-vector p for public and private goods, respectively.

We assume that each married person has her or his own preferences over the allocation of family resources. The most general version of the model would consider utilities of the form $U^a(Q, q^1, \ldots, q^K)$, implying that a is concerned directly with all members' consumptions. Here, however, tractability requires additional structure. In what follows, we therefore assume that preferences are of the *caring* type. That is, each individual a has a felicity function $u^a(Q, q^a)$; and a's utility takes the form:

[4] See Browning et al. (2014).

[5] Browning et al. (2008) and Lechene and Preston (2011) provide a set of necessary conditions for noncooperative models. However, whether these conditions are sufficient is not known; moreover, no general identification result has been derived so far.

$$U^a\left(Q, q^1, \ldots, q^K\right) = W^a\left(u^1\left(Q, q^1\right), \ldots, u^K\left(Q, q^K\right)\right), \tag{16.1}$$

where $W^a(.,.)$ is a monotone increasing function. The weak separability of these "social" preferences represents an important moral principle; a is indifferent between bundles (q^b, Q) that b consumes whenever b is indifferent. In this sense, caring is distinguished from paternalism. Caring rules out direct externalities between members because a's evaluation of her private consumption q^a does not depend directly on the private goods that b consumes.

Lastly, a particular but widely used version of caring is *egotistic* preferences, whereby members only care about their own (private and public) consumption; then individual preferences can be represented by felicities (i.e., utilities of the form $u^a(Q, q^a)$).[6] Note that such egotistic preferences *for consumption* do not exclude noneconomic aspects, such as love, companionship, or others. That is, a person's utility may be affected by the *presence* of other persons, but not by their *consumption*. Technically, the "true" preferences are of the form $F^a(u^a(Q, q^a))$, where F^a may depend on marital status and on the spouse's characteristics. Note that the F^as will typically play a crucial role in the decision to marry and in the choice of a partner. However, it is irrelevant for the characterization of married individuals' preferences over consumption bundles.

Efficiency has a simple translation; namely, the household behaves as if it was maximizing a weighted sum of utilities of its members. Technically, the program is thus (assuming egotistic preferences):

$$\max_{(Q, q^1, \ldots, q^K)} \sum_a \mu^a u^a\left(Q, q^a\right) \tag{P}$$

under the budget constraint:

$$\sum_i P_i Q_i + \sum_j p_j \left(q_j^1 + \cdots + q_j^K\right) = y^1 + \cdots + y^K = y$$

where y^a denotes a's (nonlabor) income. Here, μ^a is the *Pareto weight* of member a; one may, for instance, adopt the normalization $\sum_a \mu^a = 1$. In the particular case where μ^a is constant, the program above describes a *unitary* model, whereas household behavior is described by the maximization of some (price-independent) utility. In general, however, μ^a may vary with prices and individual incomes; the maximand in (P) is therefore price dependent, and we are not in a unitary framework in general.

This program can readily be extended to caring preferences; one must simply replace $u^a(Q, q^a)$ with $W^a(u^1(Q, q^1), \ldots, u^K(Q, q^K))$ in (P). In what follows, however, (P) plays a

[6] Throughout the chapter, we assume, for convenience, that utility functions $u^a(\cdot)$, $a = 1, K$ are continuously differentiable and strictly quasi-concave.

very special role, because any allocation that is efficient for caring preferences must be efficient for the underlying, egotistic felicities, as stated by the following result:

Proposition 1

Assume that some allocation is Pareto efficient for the caring utilities $W^1, \ldots, W^K$. Then it solves (P) for some $(\mu^1, \ldots, \mu^K)$.

Proof

Suppose there exists an alternative allocation that gives a larger value to u^a for all $a = 1, \ldots, K$. But then that allocation also gives a higher value to all W^as, a contradiction.

The converse is not true, because a very unequal solution to (P) may fail to be Pareto efficient for caring preferences: transferring resources from well-endowed but caring individuals to the poorly endowed ones may be Pareto improving. Still, any property of the solutions to a program of the form (P) must be satisfied by any Pareto-efficient allocation with caring preferences.

A major advantage of the formulation (P) is that the Pareto weights have a natural interpretation in terms of respective decision powers. The notion of "power" in households may be difficult to define formally, even in a simplified framework such as ours. Still, it seems natural to expect that when two people bargain, a person's gain increases with the person's power. This somewhat hazy notion is captured very effectively by the Pareto weights. Clearly, if μ^a in (P) is zero, then a has no say on the final allocation, whereas if μ^a is large, then a effectively gets her way. A key property of (P) is precisely that increasing μ^a will result in a move along the Pareto frontier, in the direction of higher utility for a. If we restrict ourselves to economic considerations, we may thus consider that the Pareto weight μ^a reflects a's power, in the sense that a larger μ^a corresponds to more power (and better outcomes) being enjoyed by a.

If $\left(\overline{Q}(p, P, y), \overline{q}^1(p, P, y), \ldots, \overline{q}^K(p, P, y)\right)$ denotes the solution to (P), we define the *collective indirect utility* of a as the utility reached by a at the end of the decision process; formally:

$$V^a(p, P, y) = u^a\left(\overline{Q}(p, P, y), \overline{q}^a(p, P, y)\right)$$

Note that, unlike the unitary setting, in the collective framework a member's collective indirect utility depends not only on the member's preferences but also on the decision process (hence the adjective "collective"). This notion is crucial for welfare analysis, as we shall see below.

Finally, an important concept is the notion of "distribution factors." A distribution factor is any variable that (i) does not affect preferences or the budget constraint, but (ii) may influence the decision process, therefore the Pareto weights. Think, for instance, of a bargaining model in which the agents' respective threat points may vary. A change in the threat point of one member will typically influence the outcome of the bargaining process, even if the household's budget constraint is unaffected. In particular, several tests of household behavior consider the income pooling property. The basic intuition is straightforward: in a unitary framework, whereby households behave like single decision makers (and maximize a unique, income-independent utility), only total household income should matter. Individual contributions to total income have no influence on behavior: they are pooled in the right-hand side of the household's budget constraint. For instance, paying a benefit to the wife rather than to the husband cannot possibly impact the household's demand. As we will see later, this property has been repeatedly rejected by the data. The most natural interpretation for such rejections (although not the only one) is that individual incomes may impact the decision process (in addition to their aggregate contribution to the budget constraint). Technically, if $(y^1,\ldots,y^K)$ is the vector of individual incomes and $y=\sum_a y^a$, whereas total income y is *not* a distribution factor (it enters the budget constraints), the $(K-1)$ ratios $y^1/y,\ldots,y^{K-1}/y$ are.[7] Of course, such a setting by no means implies that each individual consumes exactly his or her income. On the contrary, empirical evidence strongly suggests that transfers between family members are paramount. Whether these transfers are progressive or regressive, that is, whether they increase or decrease intrahousehold inequality, is in the end an empirical question. Whether it can be answered ultimately depends on the extent to which these transfers can be either observed or identified, an issue to which the end of this survey is dedicated.

In what follows, the vector of distribution factors will be denoted $z=(z_1,\ldots,z_S)$; Pareto weights and collective indirect utilities, therefore, have the general form $\mu^a(p,P,y,z)$ and $V^a(p,P,y,z)$.

16.3. MODELING HOUSEHOLD BEHAVIOR: THE COLLECTIVE MODEL

16.3.1 Private Goods Only: The Sharing Rule

We first consider a special case in which all commodities are privately consumed. Then the household can be considered as a small economy without externalities or public goods. From the second welfare theorem, any Pareto efficient allocation can be decentralized by adequate transfers; formally, we have the following result:

Proposition 2

Assume an allocation $(\bar{q}^1,\ldots,\bar{q}^K)$ is Pareto efficient. There then exist K nonnegative functions $(\rho^1,\ldots,\rho^K)$ of prices, total income, and distribution factors, with $\sum_k \rho^k(p,y,z)=y$, such that agent a solves

$$\max_{q^a} u^a(q^a) \tag{D}$$

[7] In practice, distribution factors must also be *uncorrelated* with unobserved components of preferences, which, in the case of individual incomes, can generate subtle exogeneity problems. See Browning et al. (2014) for a detailed discussion.

under the budget constraint

$$\sum_{i=1}^{n} p_i q_i^a = \rho^a.$$

Conversely, for any nonnegative functions $(\rho^1, \ldots, \rho^K)$, such that $\sum_k \rho_k(p, y, z) = y$, an allocation that solves (D) for all a is Pareto efficient.

In words, in a private-goods setting, any efficient decision can be described as a two-stage process. In the first stage, agents jointly decide on the allocation of household aggregate income y between agents (and agent a gets ρ^a); in the second stage, agents freely spend the shares they have received. The decision process (bargaining, for instance) takes place in the first stage; its outcome is given by the functions $(\rho^1, \ldots, \rho^K)$, which are called the *sharing rule* of the household. From a welfare perspective, there exists a one-to-one, increasing correspondence between Pareto weights and the sharing rule, at least when the Pareto set is strictly convex: when prices and incomes are constant, increasing the weight of one individual (keeping the other weights unchanged) always results in a larger share for that individual. The converse is also true. Finally, the collective, indirect utility takes a simple form, namely

$$V^a(p, y) = v^a(p, \rho^a(p, y)),$$

where v^a is the standard, indirect utility of agent a. We therefore have the following result:

Proposition 3

When all commodities are privately consumed, then for any given price vector there exists a one-to-one correspondence between the sharing rule and the indirect utility.

In particular, a member's collective indirect utility can be directly computed from the knowledge of that person's preferences and sharing rule; given the preferences, the sharing rule is a sufficient statistic for the entire decision process.

Regarding the issue of intrahousehold inequality, the key remark is that the sharing rule contains all the information required: Because all agents face the same prices, the sharing rule fully summarizes intrahousehold allocation of resources. As such, it is directly relevant for intrahousehold inequality. Specifically, let $I(y_1, \ldots, y_n)$ be some inequality index (as a function of individual incomes). Then the intrahousehold index of inequality is

$$I_I(p, y) = I\left(\rho^1(p, y), \ldots, \rho^K(p, y)\right).$$

16.3.2 Public and Private Commodities

Convenient as the previous notion may be, it still relies on a strong assumption, namely that all commodities are privately consumed. Relaxing this assumption is obviously necessary, if only because the existence of public consumption is one of the motives of household formation.

Different notions have been considered in the literature. The notion of *conditional sharing rule* (CSR), initially introduced by Blundell et al. (2005), refers to a two-stage process, whereby in stage one the household decides the consumption of public goods and the distribution of remaining income between members. In stage two all members spend their allotted amount on private consumptions to maximize individual utility *conditional on* the level of public consumption decided in stage one. As before, any efficient decision can be represented as stemming from a two-stage process of that type. The converse, however, is not true: for any given level of public consumptions, almost all CSRs lead to inefficient allocations. Moreover, the monotonic relationship between sharing rule and Pareto weights is lost. In particular, increasing a's weight does not necessarily result in a larger value for a's CSR; the intuition being that more weight to one agent may result in a different allocation of public expenditures, which may or may not result in an increase in the agent's private consumption. Lastly, and more importantly for our purpose, the CSR may give a biased estimate of intrahousehold inequality, because it simply disregards public consumption. That this pattern could be problematic is easy to see. Assume that one spouse (say the wife) cares a lot for a public good, whereas her husband cares very little. If the structure of household demand entails a significant fraction of expenditures being devoted to that public good, one can expect this pattern to have an impact on any inequality measure within the household. Disregarding public consumption altogether is therefore not an adequate approach.

A second approach relies on an old result in public economics, stating that in the presence of public goods, any efficient allocation can be decentralized using personal (or Lindahl) prices for the public good. This result establishes a nice duality between private and public goods: for the former, agents face identical prices and purchase different quantities (the sum of which is the household's aggregate demand), whereas for the latter the quantity is the same for all but prices are individual specific (and add up to market prices).[8] Again, the household behaves as if it was using a two-stage process. In stage one the household chooses a vector of individual prices for the public goods and an allocation of total income between members; in stage two members all spend their income on private and public consumptions under a budget constraint entailing their Lindahl prices. Formally, member a solves

$$\max_{q^a} u^a(Q, q^a) \tag{DP}$$

under the budget constraint

[8] See Chiappori and Ekeland (2009b) for a general presentation. For applications, see for instance Donni (2009) and Cherchye et al. (2007, 2009) for a revealed preferences perspective.

$$\sum_{i=1}^{n} p_i q_i^a + \sum_{j=1}^{N} P_j^a Q_j^a = \rho^{*a}$$

where P_j^a is the Lindahl price of good j for agent a. The vector $\rho^* = (\rho^{*1}, \ldots, \rho^{*K})$, with $\sum_a \rho^* a = y$, defines a generalized sharing rule (GSR).

From an inequality perspective, this notion raises interesting issues. One could choose to adopt ρ^* as a description of intrahousehold inequality; indeed, agents now maximize utility under a budget constraint in which ρ^* describes available income. In particular, ρ^* is a much better indicator of the distribution of resources than the CSR $\widetilde{\rho}$, because it takes into account both private and public consumptions.

However, the welfare of agent a is *not* fully described by ρ^{*a}; one also needs to know the vector P^a of a's personal prices. Technically, the collective indirect utility of a is

$$V^a(p, P, y, z) = v^a(p, P^a, \rho^{*a}(p, P, y, z)),$$

which depends on both $\rho^* a$ and P^a. This implies that the sole knowledge of the GSR is not sufficient to recover the welfare level reached by a given agent, even if her preferences are known; indeed, one also needs to know the prices, which depend on *all* preferences. In particular, we believe that the level of inequality within the household cannot be analyzed from the sole knowledge of the GSR. Agents now face different personal prices, and this should be taken into account. Of course, this conclusion was expected; it simply reflects a basic but crucial insight, namely that if agents "care differently" about the public goods (as indicated by personal prices, which reflect individual marginal willingness to pay), then variations in the quantity of these public goods have an impact on intrahousehold inequality.

Finally, Chiappori and Meghir (2014) have recently proposed the concept of Money Metric Welfare Index (MMWI). Formally, the MMWI of agent a, $m^a(p, P, y, z)$, is defined by

$$v^a(p, P, m^a(p, P, y, z)) = V^a(p, P, y, z).$$

Equivalently, if c^a denotes the expenditure function of agent a, then

$$m^a(p, P, y, z) = c^a(p, P, V^a(p, P, y, z)).$$

In words, m^a is the monetary amount that agent a would need to reach the utility-level $V^a(p, P, y)$, *if she was to pay the full price of each public good* (i.e., if she faced the price vector P instead of the personalized prices P^a). The basic intuition is simple enough. The index is defined as the monetary amount that would be needed to reach the same utility level at some reference prices. A natural benchmark is to use the current market price for all goods, private and public. In particular, there exists a direct relationship between the MMWI and the standard notion of *equivalent income*,[9] although to the best of our

[9] See for instance Fleurbaey et al. (2014).

knowledge, equivalent income has mostly been applied so far to private goods.[10] Both approaches rely on the notion that referring to a common price vector can facilitate interpersonal comparisons of welfare.

Unlike the GSR, the MMWI fully characterizes the utility level reached by the agent. That is, knowing an agent's preferences, there is a one-to-one relationship between her utility and her MMWI, and this relationship does *not* depend on the partner's characteristics. In the pure private goods case, the MMWI coincides with the sharing rule; it generalizes this notion to a general setting without losing its main advantage, namely the one-to-one relationship with welfare. Finally, it can readily be extended to allow for labor supply and domestic production; the reader is referred to Chiappori and Meghir (2014) for a detailed presentation.

16.3.3 An Example

The previous concepts can be illustrated on a very simple example, borrowed from Chiappori and Meghir (2014). Assume two agents a and b, two commodities—one private q, one public Q—and Cobb–Douglas preferences:

$$u^a = \frac{1}{1+\alpha}\log q^a + \frac{\alpha}{1+\alpha}\log Q$$

$$u^b = \frac{1}{1+\beta}\log q^b + \frac{\beta}{1+\beta}\log Q$$

corresponding to the indirect utilities

$$v^a = \log y - \frac{\alpha}{1+\alpha}\log P - \log(1+\alpha) + \frac{\alpha}{1+\alpha}\log \alpha$$

$$v^b = \log y - \frac{\beta}{1+\beta}\log P - \log(1+\beta) + \frac{\beta}{1+\beta}\log \beta.$$

Let μ be b's Pareto weight; then the couple's consumption is given by

$$q^a = \frac{1}{(1+\alpha)(1+\mu)}y, \quad q^b = \frac{\mu}{(1+\beta)(1+\mu)}y$$

$$\text{and } Q = \frac{\alpha(1+\beta)+\mu\beta(1+\alpha)}{(1+\alpha)(1+\beta)(1+\mu)}\frac{y}{P},$$

generating utilities equal to

$$V^a = \log y - \frac{\alpha}{1+\alpha}\log P - \log((1+\alpha)(1+\mu)) + \frac{\alpha}{1+\alpha}\log\left(\frac{\alpha(1+\beta)+\mu\beta(1+\alpha)}{1+\beta}\right)$$

[10] See, however, Hammond (1995) and Fleurbaey and Gaulier (2009).

$$V^b = \log y - \frac{\beta}{1+\beta}\log P$$
$$- \log(1+\beta)(1+\mu) + \frac{1}{1+\beta}\log \mu + \frac{\beta}{1+\beta}\log\left(\frac{\alpha(1+\beta)+\mu\beta(1+\alpha)}{1+\alpha}\right).$$

In this context, straightforward calculations allow to see that

1. The CSR coincides with private consumption

$$\widetilde{\rho}^a = \frac{1}{(1+\alpha)(1+\mu)}y, \quad \widetilde{\rho}^b = \frac{\mu}{(1+\beta)(1+\mu)}y$$

2. Lindahl prices are

$$P^a = \frac{\alpha(1+\beta)}{\alpha(1+\beta)+\mu\beta(1+\alpha)}P$$

$$P^b = \frac{\mu\beta(1+\alpha)}{\alpha(1+\beta)+\mu\beta(1+\alpha)}P$$

and the GSR is

$$\rho^{*a} = \frac{y}{1+\mu}$$

$$\rho^{*b} = \frac{\mu y}{1+\mu}$$

3. The two MMWIs are given by

$$m^a = \left(\frac{\alpha(1+\beta)+\mu\beta(1+\alpha)}{\alpha(1+\beta)}\right)^{\frac{\alpha}{1+\alpha}}\frac{y}{1+\mu} = \left(\frac{\alpha(1+\beta)+\mu\beta(1+\alpha)}{\alpha(1+\beta)}\right)^{\frac{\alpha}{1+\alpha}}\rho^{*a}$$

$$m^b = \left(\frac{\alpha(1+\beta)+\mu\beta(1+\alpha)}{\mu\beta(1+\alpha)}\right)^{\frac{\beta}{1+\beta}}\frac{\mu y}{1+\mu} = \left(\frac{\alpha(1+\beta)+\mu\beta(1+\alpha)}{\mu\beta(1+\alpha)}\right)^{\frac{\beta}{1+\beta}}\rho^{*b}.$$

Assume now that $\mu=1$, but agents have different preferences for the public good. For instance, $\alpha=2$ while $\beta=0.5$, implying that the wife (or husband) puts two-thirds of the weight on the public (private) consumption. In this setting, we can analyze intrahousehold inequality using three possible indicators.

1. If we concentrate on private consumption (or equivalently on the CSR), we find that

$$\widetilde{\rho}^a = \frac{1}{6}y, \quad \widetilde{\rho}^b = \frac{1}{3}y$$

and we conclude that member b is much better off than a.

2. This conclusion is clearly unsatisfactory, because it disregards the fact that half the budget is spent on the public good, which benefits a more than b. Indeed, the GSR is
$$\rho^{*a} = \frac{y}{2} = \rho^{*b}$$
and we conclude that for this indicator, the household is perfectly equal: the benefits of public expenditures exactly compensate differences in private consumptions.
3. The later conclusion is, however, too optimistic, as it omits the fact that a "pays" twice as much for the public good than b does (here, $P^a = \frac{2}{3}P$ while $P^b = \frac{1}{3}P$). Taking this last aspect into account, the respective MMWIs are
$$m^a = 0.655y, \quad m^b = 0.72y$$

Again, b is better off than a (although by much less than with the first measure). In addition, one may note that
$$m^a + m^b = 1.375y.$$
Individual MMWIs add up to more than total income, reflecting the gain generated by the publicness of one commodity.

16.3.4 Domestic Production

Finally, the previous analysis can readily be extended to domestic production. Here, we only consider the case where all commodities are privately consumed; for a more general presentation along similar lines, the reader is referred to Chiappori and Meghir (2014). The household production technology is thus described by a production function that gives the possible vector of outputs $q = f(x, \tau)$, that can be produced given a vector of market purchases x and the time $\tau = (\tau^a, a = 1, K)$ spent in household production by each of the members.

We first disregard the time spent by each member on domestic production. This setting is thus identical to the general model of household production of Browning et al. (2013).[11] Pareto efficiency translates into the program
$$\max \sum \mu^a u^a (q^a)$$
$$\sum_a q_i^a = f_i (x^i),$$
$$p' \left(\sum_i x^i \right) = y,$$
where

[11] For empirical applications, these authors use a linear technology a la Barten.

$$q^a = \left(q_i^a\right), \quad i = 1, n$$

$$x^i = \left(x_j^i\right), \quad j = 1, k.$$

As before, this program can be decentralized, although decentralization now requires specific (shadow) prices for the produced goods. Specifically, let η_i, λ be the respective Lagrange multipliers of the production constraints in (16.2), and define

$$\pi_i = \frac{\eta_i}{\lambda}.$$

Let $((q^a*), a = 1, \ldots, K, x*)$ denote the solutions, and define the sharing rule by

$$\rho^a = \pi' q^{a*}.$$

Then the program is equivalent to a two-stage process, in which q^a* solves

$$\max u^a(q^a)$$

under the budget constraint

$$\pi' q^a = \rho^a$$

and $x*$ solves the profit maximization problem

$$\max \sum_i \pi_i f_i\left(x^i\right) - \sum_{i,j} p_j x_j^i,$$

or, equivalently, the cost minimization problem

$$\min p' x$$

$$f(x) = \sum_a q^{a*}.$$

In that case, again, individual welfare is adequately measured by the sharing rule.

Extending this model to domestic labor supply is straightforward. The Pareto program is now

$$\max \sum \mu^a u^a(q^a, L^a)$$

$$\sum_a q_i^a = f_i\left(x^i, \tau_i\right)$$

$$p'\left(\sum_i x^i\right) + \sum_a w_a \left(L^a + \sum_i \tau_i^a\right) = y + \sum_a w_a T = Y,$$

where

$$\tau_i = \left(\tau_i^a\right), \quad a = 1, K.$$

Prices for internally produced goods are defined as before. The sharing rule is now

$$\rho^a = \pi' q^{a*} + w_a L^{a*}, \quad a = 1, K,$$

where L^a* denotes a's optimal leisure. The program can be decentralized as follows. For each a, (q^a*, L^a*) solve

$$\max u^a(q^a, L^a),$$
$$\pi' q^a + w_a L^a = \rho^a,$$

and $x*, \tau^a*$ solves

$$\max \sum_i \pi_i f_i(x^i, \tau_i) - \sum_{i,j} p_j x_j^i - \sum_{i,a} w_a \tau_i^a.$$

or equivalently,

$$\min \sum_{i,j} p_j x_j^i + \sum_{i,a} w_a \tau_i^a \tag{16.2}$$

under

$$f_i(x^i, \tau_i^a) = \sum_a q_i^{a*}, \quad i = 1, n.$$

In practice, several variants of this basic framework can be considered, depending on whether the internally produced goods are marketable, and whether market labor supplies are available via an interior or a corner solution. These technical issues are not without importance. For instance, a standard issue in family economics is whether a change in the respective powers of the various members has an impact on the intrahousehold allocation of domestic work. In the model just described, if the produced commodities are marketable and all individuals work on the market, then the πs and the ws must coincide with market prices and wages; they are therefore exogenous, and individual, domestic labor supplies are fully defined by the program (16.2), which does not depend on Pareto weights. We conclude that, in that case, powers have no impact on domestic work, which is fully determined by efficiency considerations. Clearly, this argument must be modified when either the πs or the ws are endogenous (as will be the case if, respectively, the commodity is not marketable or a person does not participate in the labor market). The reader is referred to Browning et al. (2014) for a precise discussion, as well as to the Chapter 12 in this handbook.

16.4. THE DETERMINANTS OF INTRAHOUSEHOLD ALLOCATION

The second task assigned to theory is to explain the allocation of powers, hence of resources, within the household. As such, it must address issues related to household formation and dissolution, as well as the interaction between the household and its environment, that is, which external factors may impact the intrahousehold decision process. In what follows, we concentrate on two types of approaches, respectively based on cooperative bargaining and matching or search theory. In a sense, this distinction reflects the classic dichotomy between partial and general equilibrium. Bargaining models analyze, for a given household, how the particular situation of each member may affect the household decision; much emphasis is put on individual "threat points," generally considered as exogenous. Matching and search models, on the other hand, describe a global equilibrium on the "market for marriage" as a whole. Although the decision process may in some cases entail bargaining (in search models, or in matching with a finite set of agents), the crucial distinction is that the threat points are now *endogenous*—their determination is part of the equilibrium conditions.

16.4.1 Bargaining Models

Any bargaining model requires a specific setting: In addition to the framework described above (K agents, with specific utility functions), one has to define a *threat point* T^a for each individual a. Intuitively, a person's threat point describes the utility level this person could reach in the absence of an agreement with the partner. Typically, bargaining models assume that the outcome of the decision process is Pareto efficient and individually rational, in the sense that individuals never receive less than their threat point. Bargaining theory is used to determine how the threat points influence the location of the chosen point on the Pareto frontier. Clearly, if the point $T=(T^1,\ldots,T^K)$ is outside of the Pareto set, then no agreement can be reached, because at least one member would lose by agreeing. However, if T belongs to the interior of the Pareto set so that all agents can gain from the relationship, the model picks a particular point on the Pareto utility frontier. Note that the crucial role played by threat points—a common feature of all bargaining models—has a very natural interpretation in terms of distribution factors. Indeed, *any variable that is relevant for threat points only is a potential distribution factor*. For example, the nature of divorce settlements, the generosity of single-parent benefits, or the probability of remarriage do not directly change a household's budget constraint (as long as it does not dissolve), but may affect the respective threat points of individuals within it. Then bargaining theory implies that they will influence the intrahousehold distribution of power in households and, ultimately, household behavior. Equivalently, one could say that these variables are distribution factors that affect the Pareto weights.

In practice, models based on bargaining must make a number of basic choices. One is the bargaining concept to be used. Whereas most studies refer to Nash bargaining, some either adopt Kalai–Smorodinski or refer to a noncooperative bargaining model. Second, one must choose a relevant threat point. This part is crucial; indeed, a result due to Chiappori et al. (2012) states that *any* Pareto efficient allocation can be derived as the

Nash bargaining solution for an *ad hoc* definition of the threat points. Hence any additional information provided by the bargaining concepts (besides the sole efficiency assumption) must come from specific hypotheses on the threat points, that is, on what is meant by the sentence "no agreement is reached." Several ideas have been used in the literature. One is to refer to divorce as the "no agreement" situation. Then the threat point is defined as the maximum utility a person could reach after divorce. Such an idea seems well adapted when one is interested, say, in the effects of laws governing divorce on intrahousehold allocation. It is probably less natural when minor decisions are at stake: Choosing who will walk the dog, for example, is unlikely to involve threats of divorce.[12] Another interesting illustration would concern public policies that affect single parents, or the guaranteed employment programs that exist in some Indian states. Haddad and Kanbur (1992) convincingly argue that the main impact of the program was to change the opportunities available to the wife outside of marriage.

A second idea relies on the presence of public goods and the fact that noncooperative behavior typically leads to inefficient outcomes. The idea, then, is to take the noncooperative outcome as the threat point: In the absence of an agreement, both members provide the public good(s) egotistically, not taking into account the impact of their decision on the other member's welfare. This version captures the idea that the person who would suffer more from this lack of cooperation (the person who has the higher valuation for the public good) is likely to be more willing to compromise in order to reach an agreement. A variant, proposed by Lundberg and Pollak (1993), is based on the notion of "separate spheres." The idea is that each partner is assigned a set of public goods to which they alone can contribute; this is their "sphere" of responsibility or expertise. These spheres are determined by social norms. Then the threats consist of continued marriage in which the partners act noncooperatively and each chooses independently the level of public goods under their domain.

Finally, it must be reminded that assumptions on threat points tend to be strong, not grounded on strong theoretical arguments, and often not independently testable. This suggests that models based on bargaining should be used parsimoniously and with care.

16.4.2 Equilibrium Models

Alternatively, one can consider the "market for marriage" as a whole from a general perspective. Two types of models can be found in the literature, that make opposite assumptions on the role of frictions in the matching game. Specifically, models based on matching (with transferable or imperfectly transferable utility, TU and ITU, respectively) assume away frictions and consider perfectly smooth markets, while models based on

[12] An additional difficulty is empirical. The estimation of utility in case of divorce is delicate, because most data sets allow us to estimate (at best) an ordinal representation of preferences, whereas Nash bargaining requires a cardinal representation. See Chiappori (1991).

search emphasize the importance of frictions in the emergence of marital patterns. While matching and search-based approaches use different technologies, their scope and outcomes are largely similar for what we are concerned with here. In what follows, for the sake of brevity, we therefore concentrate on matching models. Moreover, we only discuss models based on TU. The nontransferable utility framework, which assumes away any transfer between members, is not relevant here. And although more general approaches based on ITU have recently been developed (see Chiappori, 2012), the distinction between TU and ITU can basically be disregarded for our current discussion.

Consider the two populations of men and women: Each individual is defined by a vector of characteristics, denoted $x \in X$ for women and $y \in Y$ for men. Both sets are endowed with a finite measure, denoted μ_X and μ_Y, respectively. When matched, Mrs. x and Mr. y jointly generate a surplus $s(x, y)$, which can be derived from a more structural framework (e.g., a collective model). A *matching* is defined by (i) a measure μ on the set $X \times Y$, the marginals of which coincide with μ_X and μ_Y, and (ii) two functions $u(x)$ and $v(y)$ such that $u(x) + v(y) = s(x, y)$ on the support of μ. Intuitively, the measure μ defines who marries whom, whereas the functions determine how the surplus is divided within couples who are matched with positive probability: She gets $u(x)$, he gets $v(y)$. A matching is *stable* if (i) no married person would prefer being single and (ii) no pair of currently unmarried persons would both prefer forming a new couple. Technically, this is equivalent to

$$u(x) + v(y) \geq s(x, y) \quad \forall (x, y). \tag{16.3}$$

The functions $u(x)$ and $v(y)$ are crucial, inasmuch as they fully determine the intrahousehold inequality. The key feature of matching models is that these functions are *endogenous*. They are determined (or constrained) as part of the equilibrium, and depend on the whole matching game structure; in particular, the allocation within any given couple depends on the entire distribution of characteristics in the two populations. In that sense, the model does provide an endogenous determination of intrahousehold inequality. Note, however, that in this abstract presentation, their exact interpretation is undetermined; depending on the framework, $u(x)$ can be a monetary amount, the consumption of some commodity or the utility generated by the consumption of bundles of private and public commodities. For instance, the simple framework used by Chiappori and Weiss (2007) consider an economy with two commodities, one private and one public within the household, and agents with Cobb–Douglas preferences $u^a = q^a Q$; x and y are one-dimensional and denote male and female income. In this TU framework, any efficient allocation maximizes the *sum* of utilities; that is, a (x, y) couple solves

$$\max_{q^1, q^2, Q} \left(q^1 + q^2\right) Q \quad \text{under} \quad q^1 + q^2 + Q = x + y \tag{16.4}$$

and the surplus $s(x, y)$ is the value of this program, namely $(x + y)^2/4$. Here, $u(x)$ and $v(y)$ are utilities, although there exists a one-to-one correspondence between utilities and transfers (because $Q = (x + y)/2$, we have that $q^1 = 2u(x)/(x + y), q^2 = 2v(y)/(x + y)$).

From a mathematical point of view, a basic result states that, if a matching is stable, then the corresponding measure maximizes total surplus over the set of measures whose marginals coincide with μ_X and μ_Y. That is, the measure μ must solve

$$\max_{\mu} \int_{X \times Y} s(x, y) \mathrm{d}\mu(x, y) \tag{16.5}$$

under the marginal conditions. This maximization problem is linear in its unknown μ. Therefore, it admits a dual, which can be written as

$$\min_{u, v} \int_X u(x) \mathrm{d}\mu_X(x) + \int_Y v(y) \mathrm{d}\mu_Y(y) \tag{16.6}$$

under the constraints

$$u(x) + v(y) \geq s(x, y) \quad \forall (x, y) \tag{16.7}$$

Here, functions u and v are the dual variables of the program. But, crucially, they can be interpreted as describing the utility reached by each individual at optimal matching. In particular, they define the allocation of surplus between (matched) spouses. Note that conditions of the dual program (16.7) are exactly the stability conditions (16.3).

From the standard duality results, a solution to the dual exists if and only if the primal has a solution, and the values are then the same. It follows that the existence of a stable match (that is, of functions u and v satisfying (16.3)) boils down to the existence of a solution to the linear maximization problem (16.5). This allows us to establish existence under very general conditions; see, for instance, Chiappori et al. (2010).

Regarding uniqueness, if the sets X and Y are finite, then the u s and vs are not pinned down, although the equilibrium conditions generate constraints. However, with continuous, atomless populations, the functions are in general fully determined by the equilibrium conditions. The intuition is straightforward: in the continuous case, each individual has almost perfect substitutes, and (local) competition determines exactly the surplus sharing that must exist at equilibrium. Finally, stochastic versions of these models can be considered, in which some of the individual characteristics are unobserved (to the econometrician); see, for instance, the recent survey by Chiappori and Salanié (2014).

16.5. IDENTIFICATION

Whereas the conceptual tools just presented help clarify some of the issues involved, their empirical content must be very carefully considered. As stated previously, there is no point in putting much emphasis on a concept that cannot possibly be identified from existing data. This section summarizes the main results obtained on this issue over the last two decades; for a more detailed presentation, the reader is referred to Chiappori and Ekeland (2009a).

We divide the presentation into three subsections. One considers the "pure" identification problem. Assume that the entire demand function of a household can be observed; what can be recovered from such data (and such data only)? Next, we introduce additional identifying assumptions. Broadly speaking, these postulate a relationship between an individual's preferences as a single person and as being part of a household; in other words, we admit that some information about spouses' utilities can be derived from the observation of the behavior of single persons. Lastly, we introduce a general, market-wide perspective and ask whether (and how) equilibrium conditions on the marriage market can help identify the intrahousehold allocation process.

16.5.1 "Pure" Identification in the Collective Model

Identification issues in the collective model have been extensively studied during the recent years; the interested reader is referred to Chiappori and Ekeland (2009a,b) for an exhaustive presentation. In what follows, we briefly summarize some key findings.

We start with the basic framework described above, assuming egotistic preferences of the type $u^a(Q, q^a)$; also, for the sake of brevity, we assume only two persons (spouses) in the household, although the generalization to any number is straightforward. In what follows, we assume that we observe the household's "aggregate" demand, that is, the vector $(q, Q) \in \mathbb{R}^{n+N}$ (where $q_i = \sum_a q_i^a, i = 1, \ldots, n$) as a function of prices (p, P) and total income y, plus possibly a vector of distribution factors z. Remember that the *collective indirect utility* of agent a is defined as the utility level a will reach at the end of the decision process, as a function of (p, P, y, z).

16.5.1.1 Main Identification Result

Assume, first, that we observe the demand function of some household. This demand is aggregated at the household level. This means what we observe is the household's total demand for any private commodity, together with its demand for public goods. However, in general, we are not able to observe the internal allocation of the private goods between household members. When is this information sufficient to recover the underlying structure, that is, the preferences and the decision process (as summarized by the Pareto weights)?

A first answer is provided as a result from Chiappori and Ekeland (2009a). It states that generically, all that is needed is one exclusion restriction per agent. In other words, for any agent a, there should be some commodity that a does *not* consume (and which does not enter a's egoistic utility). Then the *local* knowledge of the household demand allows us to exactly (locally) identify each agent's collective indirect utility, irrespective of the number of private and public goods. Formally, this is stated as

Theorem 1

(*Chiappori and Ekeland, 2009a*) *Assume* $N + n \geq 4$. *Consider a point* $(\overline{p}, \overline{P}, \overline{y})$ *such that the CSR satisfies the condition*

$$\frac{\partial \rho^a}{\partial y}\left(\overline{p}, \overline{Q}, \overline{y}\right) \neq 0, \quad a = 1, 2,$$

where $\overline{Q} = Q(\overline{p}, \overline{P}, \overline{y})$. *Assume that for each member, there exists at least one good not consumed by this member (but consumed by the other). Then generically there exists an open neighborhood of* $(\overline{p}, \overline{P}, \overline{y})$ *on which the indirect collective utility of each member is exactly (ordinally) identifiable from household demand. For any cardinalization of indirect collective utilities, the Pareto weights are exactly identifiable.*

For a precise proof, see Chiappori and Ekeland (2009a) Proposition 7 on page 781. The underlying intuition is that if commodity i is not consumed by agent y, then any impact of its price on that agent's behavior can only operate through the decision process—the Pareto weights. The resulting conditions, which are reminiscent of separability restrictions in standard consumer theory, are sufficient in general to fully recover the (ordinal) indirect collective utility of each member, as well as, for any choice of cardinalization, the corresponding Pareto weights.

The specific nature of the identification result can be simply illustrated on a Cobb–Douglas example, as described later on. Before considering it, a few remarks are in order. First, the identification result stated in Theorem 1 is only local. This is important because additional constraints of a global nature (such as nonnegativity restrictions on consumption), which are not considered in this result, typically provide additional identification power; a precise illustration will be given below. Second, the result does not require distribution factors. Again, the latter would allow a stronger identification result. Indeed, Chiappori and Ekeland show that, in the presence of distribution factors, the exclusivity requirement can be relaxed; one only needs either one excluded good (instead of two) or an *assignable* commodity.[13] Third, identification requires the observation of the household demand as a function of prices and income; in particular, price variations are crucial. Although this fact is not surprising (even in standard consumer theory, preferences cannot be recovered from demand without price variations) it has important empirical applications, because data entailing significant (and credibly exogenous) price variations are not easy to find. However, recent approaches relax this requirement by imposing additional structure on the decision process; they will be discussed below.

Fourth, the identification result above is only generic: it may fail to hold in particular cases, although such cases are not robust to "small variations." Quite interestingly, one of the situations in which identification does not obtain is the unitary model. To see why, consider program (P) above, and assume that the Pareto weights μ^a are all constant. For

[13] A good is assignable when it is consumed by both members, and the consumption of each member is independently observed.

one thing, we are in a unitary context: the household maximizes the sum $\sum_a \mu^a u^a(Q, q^a)$, which is a *price- and income-independent* utility. More importantly, Hicks's aggregation theorem applies. If we define U by

$$U(Q, q) = \max_{\sum_a q^a = q} \sum_a \mu^a u^a(Q, q^a) \tag{16.8}$$

then the household maximizes U under the budget constraint. By standard integration, U can be recovered from the household demand. However, this is not sufficient to identify individual preferences: there exists a continuum of different sets of individual utilities that generate the same U by (16.8). The paradox here is that the unitary model, which used to be the dominant framework for empirical works on household behavior, belongs to the small (actually nongeneric) class of frameworks for which individual welfare cannot be identified from household demand.

Lastly, it is important to note that what is identified is the *indirect collective utility* of each member. From a welfare perspective, this is the only relevant concept, because it fully characterizes the utility reached by each agent. However, the inequality measures described above require more, namely, an assessment of the intrahousehold allocation of income. We now consider to what extent the latter can be recovered from the indirect collective utility.

16.5.1.2 Private Goods and the Sharing Rule

We start with the case in which all commodities are private. In that case, the various concepts (CSR, GSR, MMWI) coincide with the sharing rule, and the collective indirect utility takes the form

$$V^a(p, y) = v^a(p, \rho^a(p, y)),$$

where, as above, v^a is a's indirect utility and ρ is the sharing rule. If we assume that the first (respectively the second) good is exclusively consumed by the second (first) agent, the collective indirect utility of each agent is identified (as always, up to some increasing transform).

16.5.1.2.1 Local Identification

A first result states that the sharing rule is *not* fully identified from the knowledge of the collective indirect utility, at least locally; identification only obtains up to an additive function of the prices of the nonexclusive goods. Formally, assume that one observes the functions $(q_1, \ldots q_n)$ of (p, y), with $p \in \mathbb{R}^n$ and

$$\begin{aligned} q_1(p, y) &= \chi_1^a(p, \rho(p, y)) \\ q_2(p, y) &= \chi_2^b(p, y - \rho(p, y)) \\ q_i(p, y) &= \chi_i^a(p, \rho(p, y)) + \chi_i^b(p, y - \rho(p, y)), \quad i = 3, \ldots, n, \end{aligned} \tag{16.9}$$

where the functions χ_i^s and ρ are unknown. Then

Proposition 4

(Chiappori and Ekeland, 2009a) Assume $n \geq 3$, and let $\left(\overline{\chi}_1^a, \ldots, \overline{\chi}_n^b, \overline{\rho}\right)$ solve for (16.9). For any other solution $(\chi_1^a, \ldots, \chi_n^b, \rho)$, there exist a $\varphi : \mathbb{R}^{n-2} \to \mathbb{R}$ such that

$$\begin{aligned} \rho(p, y) &= \overline{\rho}(p, y) + \varphi(p_3, \ldots, p_n), \\ \chi_i^a(p, \rho) &= \overline{\chi}_i^a(p, \rho - \varphi(p_3, \ldots, p_n)), \\ \chi_j^b(p, \rho) &= \overline{\chi}_j^b(p, \rho + \varphi(p_3, \ldots, p_n)). \end{aligned} \tag{16.10}$$

Moreover, overidentifying restrictions are generated.

The basic conclusion is that the sharing rule is identified up to an additive function, which cannot be pinned down unless either all commodities are assignable or individual preferences are known (for instance, from data on singles) or other (global) restrictions are used as described below. To see why, consider the simple case of three private commodities; two of these are exclusive (for members *a* and *b*, respectively), whereas the third is consumed by both. Individual consumptions of commodity 3 are not observed, and its price is taken as numeraire. In practice, we observe two demand functions q_1^a and q_2^b that satisfy

$$q_1^a(p_1, p_2, y) = \widetilde{q}^a(p_1, \rho(p_1, p_2, y)) \tag{16.11}$$

$$q_2^b(p_1, p_2, y) = \widetilde{q}^b(p_2, y - \rho(p_1, p_2, y)) \tag{16.12}$$

where $\widetilde{q}^s$ denotes the Marshallian demand by person *s*. Now, for some constant *K*, define ρ_K, u_K^a and u_K^b by

$$\begin{aligned} \rho_K(p_1, p_2, y) &= \rho(p_1, p_2, y) + K, \\ u_K^a\left(q_1^a, q_3^a\right) &= u_K^a\left(q_1^a, q_3^a - K\right), \\ u_K^b\left(q_2^b, q_3^b\right) &= u_K^b\left(q_2^b, q_3^b + K\right). \end{aligned}$$

It is easy to check that the Marshallian demands derived from ρ_K, u_K^a and u_K^b satisfy (16.11) and (16.12). The intuition is illustrated in Figure 16.2 in the case of *a*. Switching from ρ and u^a to ρ_K and u_K^a does two things. First, the sharing rule and the intercept of the budget constraint are shifted downward by *K*. Second, all indifference curves are also shifted downward by the same amount. When only demand for commodity 1 (on the horizontal axis) is observable, these models are empirically indistinguishable. Lastly, with several, nonexclusive goods, this construct is still possible, and the constant may in addition vary with nonexclusive prices in an arbitrary way.

Two remarks can be made about this result. One is that the indetermination is not welfare relevant; one can easily check that the different solutions correspond to the same collective indirect utilities for each agent. This is the paradox evoked in introduction. Unlike standard consumer theory, there is no longer an equivalence between identifying direct and indirect utilities. Indirect utilities are identified as soon as the exclusion restrictions are satisfied, but they may correspond to various, welfare-equivalent direct utilities, each of them associated with a specific sharing rule.

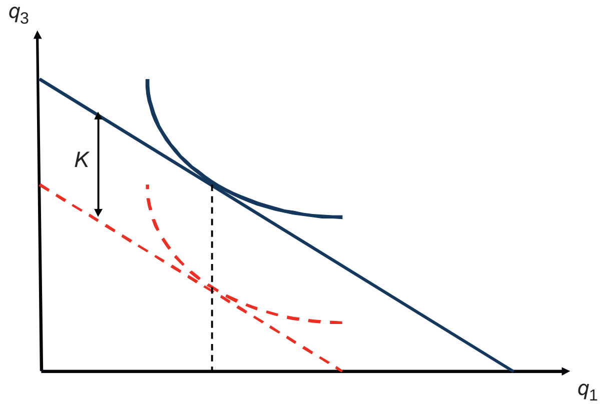


Figure 16.2 Welfare equivalence of alternative levels of the sharing rule.

16.5.1.2.2 Global Restrictions

The second remark is that the nonidentification result is only local. In particular, it disregards additional, global restrictions such as nonnegativity constraints. If these are added, then more can be identified. For instance, consider (16.10), and add the restrictions that

$$\rho(p, 0) = 0 \quad \forall p,$$

which stems from nonnegativity of consumption at very low income levels. Then φ is exactly pinned down:

$$\varphi(p_3, \ldots, p_n) = -\overline{\rho}(p, 0)$$

and additional, overidentifying restrictions are generated (e.g., $\partial\overline{\rho}(p, 0)/\partial p_i = 0$ for $i = 1, 2$).

This result should be related to recent work on the estimation of the sharing rules based on a revealed preference approach (see, for instance, Cherchye et al., 2012). Because the revealed preference approach is global by nature, it can generate bounds on the sharing rule, which can actually be quite narrow. In all cases, the global restrictions are generated at one end of the distribution of expenditures, so their use for identifying the sharing rule outside this range should be submitted to the usual caution. Still, they tend to considerably reduce the scope of the nonidentification conclusion.

16.5.1.3 Public Goods Only

We now consider the opposite polar case, in which all commodities (but the exclusive ones) are public. That is, utilities are now of the form

$$U^a(Q_1, Q_3, \ldots, Q_N) \text{ and } U^b(Q_2, Q_3, \ldots, Q_N).$$

Note that the exclusive commodities 1 and 2 can be considered as either public or private.

In that case, the collective indirect utility has a simple form, namely,

$$V^a(P, y) = U^a(Q_1, Q_3, \ldots, Q_N),$$
$$V^b(P, y) = U^b(Q_2, Q_3, \ldots, Q_N).$$

The crucial remark is that the demands for public goods (as functions of prices and total income) are empirically observed. An important consequence is that, in general, the knowledge of indirect collective utilities is equivalent to that of direct utilities. To see why, normalize y to be 1 (by homogeneity), and take a point at which the Jacobian matrix $D_P(Q_1, Q_2, \ldots, Q_N)$ is of full rank. By the implicit function theorem, we can locally invert the function, thus defining P as a function of Q; but then,

$$U^a(Q_1, Q_3, \ldots, Q_N) = V^a(P_1(Q_1, Q_2, \ldots, Q_N), \ldots, P_N(Q_1, Q_2, \ldots, Q_N), 1),$$
$$U^b(Q_2, Q_3, \ldots, Q_N) = V^b(P_1(Q_1, Q_2, \ldots, Q_N), \ldots, P_N(Q_1, Q_2, \ldots, Q_N), 1),$$

which proves identification. In addition, overidentifying restrictions are generated. In particular, we see that in this context Lindahl prices for all goods (therefore the MMWIs) are exactly identified. Somewhat paradoxically, the pure public good case appears to be the one in which identification is least problematic.

16.5.1.4 The General Case

Finally, the general case is a direct generalization of the two particular cases just described. The aforementioned exclusion restrictions guarantee identification of the collective indirect utility of each agent. Then the exact intrahousehold allocation is locally identified up to an additive function of the prices of the nonexclusive private goods. Moreover, global restrictions (e.g., nonnegativity) allow exact identification in general. The interested reader is referred to Chiappori and Ekeland (2009a,b) for detailed statements.

16.5.1.5 A Linear Expenditure System Example

The previous discussions can be illustrated on a simple example, borrowed from Chiappori and Ekeland (2009a). Consider individual preferences of the LES type:

$$U^s(q^s, Q) = \sum_{i=1}^{n} \alpha_i^s \log\left(q_i^s - c_i^s\right) + \sum_{j=n+1}^{N} \alpha_j^s \log\left(Q_j - C_j\right), \quad s = a, b,$$

where the parameters α_i^s are normalized by the condition $\sum_{i=1}^{N} \alpha_i^s = 1$ for all s, whereas the parameters c_i^s and C_j are unconstrained. Here, commodities 1 to n are private, whereas commodities $n+1$ to N are public. Also, given the LES form, it is convenient to assume that the household maximizes the weighted sum $\mu U^a + (1-\mu)U^b$, where the Pareto weight μ has the simple, linear form:

$$\mu = \mu^0 + \mu^y y + \mu^z z, \quad s = a, b.$$

16.5.1.5.1 Household Demand

The couple solve the program

$$\max\left(\mu^0+\mu^y y+\mu^z z\right)\left(\sum_{i=1}^{n}\alpha_i^a\log\left(q_i^a-c_i^a\right)+\sum_{j=n+1}^{N}\alpha_j^a\log\left(Q_j-C_j\right)\right)$$
$$+\left(1-\left(\mu^0+\mu^y y+\mu^z z\right)\right)\left(\sum_{i=1}^{n}a_i^b\log\left(q_i^b-c_i^b\right)+\sum_{j=n+1}^{N}\alpha_j^b\log\left(Q_j-C_j\right)\right)$$

under the budget constraint. Individual demands for private goods are given by

$$p_iq_i^a=p_ic_i^a+\alpha_i^a\left(\mu^0+\mu^y y+\mu^z z\right)\left(y-\sum_{i,s}p_ic_i^s-\sum_j P_jC_j\right)$$

$$p_iq_i^b=p_ic_i^b+\alpha_i^b\left[1-\left(\mu^0+\mu^y y+\mu^z z\right)\right]\left(y-\sum_{i,s}p_ic_i^s-\sum_j P_jC_j0\right)$$

generating the aggregate demand

$$p_iq_i=p_ic_i+\left[\alpha_i^a\left(\mu^0+\mu^y y+\mu^z z\right)+\alpha_i^b\left(1-\left(\mu^0+\mu^y y+\mu^z z\right)\right)\right]Y \tag{16.13}$$

and for public goods

$$P_jQ_j=P_jC_j+\left[\alpha_j^a\left(\mu^0+\mu^y y+\mu^z z\right)+\alpha_j^b\left(1-\left(\mu^0+\mu^y y+\mu^z z\right)\right)\right]Y,$$

where $c_i=c_i^a+c_i^b$ and $Y=(y-\sum_{i,s}p_ic_i^s-\sum_jP_jC_j)$. The household demand is thus a direct generalization of the standard LES, with additional quadratic terms in y^2 and cross terms in yp_i and yP_j, plus terms involving the distribution factor z.

A first remark is that c_i^a and c_i^b cannot be individually identified from group demand, because the latter only involves their sum c_i. As a consequence, the various generalizations of the sharing rule will only be identified up to one additive constant, a result mentioned earlier. Also, the constant is welfare irrelevant; indeed, the collective indirect utilities of the wife and the husband are (up to an increasing transform)

$$\begin{aligned}W^a(p,P,y,z)&=\log Y+\log\left(\mu^0+\mu^y y+\mu^z z\right)\\&\quad-\sum_i\alpha_i^a\log p_i-\sum_j\alpha_j^a\log P_j\\W^b(p,P,y,z)&=\log Y+\log\left(1-\left(\mu^0+\mu^y y+\mu^z z\right)\right)\\&\quad-\sum_i\alpha_i^b\log p_i-\sum_j\alpha_j^b\log P_j\end{aligned}$$

which does not depend on each c_i^s separately. Second, the form of aggregate demands is such that private and public goods have exactly the same structure. We, therefore, simplify our notations by defining

$$\xi_i = q_i \quad \text{for } i \leq n, \xi_i = Q_i \quad \text{for } n < i \leq N$$

and similarly

$$\gamma_i = c_i \quad \text{for } i \leq n, \gamma_i = C_i \quad \text{for } n < i \leq N,$$
$$\pi_i = p_i \quad \text{for } i \leq n, \; \pi_i = P_i \quad \text{for } n < i \leq N,$$

so that the group demand has the simple form

$$\pi_i \xi_i = \pi_i \gamma_i + \left[\alpha_i^a \left(\mu^0 + \mu^y y + \mu^z z\right) + \alpha_i^b \left(1 - \left(\mu^0 + \mu^y y + \mu^z z\right)\right)\right] Y \tag{16.14}$$

leading to collective indirect utilities of the form

$$W^a(p, P, y, z) = \log Y + \log\left(\mu^0 + \mu^y y + \mu^z z\right) - \sum_i \alpha_i^a \log \pi_i,$$
$$W^b(p, P, y, z) = \log Y + \log\left(1 - \left(\mu^0 + \mu^y y + \mu^z z\right)\right) - \sum_i \alpha_i^b \log \pi_i.$$

It is clear on this form that the distinction between private and public goods can be ignored. This illustrates an important remark: while the ex ante knowledge of the public versus the private nature of each good is necessary for the identifiability result to hold in general, for many parametric forms it is actually not needed.

16.5.1.5.2 Identifiability: The General Case

The question now is whether the empirical estimation of the form (16.14) allows us to recover the relevant parameters, namely, the α_i^s, the γ^i, and the μ^α. We start by rewriting (16.14) as

$$\pi_i \xi_i = \pi_i \gamma_i + \begin{pmatrix} \alpha_i^b + \left(\alpha_i^a - \alpha_i^b\right)\mu^0 \\ + \left(\alpha_i^a - \alpha_i^b\right)\left(\mu^y y + \mu^z z\right) \end{pmatrix} \left(y - \sum_m \pi_m \gamma^m\right) \tag{16.15}$$

The right-hand side of (16.15) can in principle be econometrically identified; we can thus recover the coefficients of the right hand side variables, i.e., y, y^2, yz, the π_m, and the products $y\pi_m$ and $z\pi_m$. For any i and any $m \neq i$, the ratio of the coefficient of y by that of π_m gives γ^m; the γ^m are therefore vastly overidentified. However, the remaining coefficients are identifiable only up to an arbitrary choice of two of them. Indeed, an empirical estimation of the right-hand side of (16.15) can only recover for each j the respective coefficients of y, y^2, and yz, that is the three expressions

$$K_y^j = \alpha_j^b + \left(\alpha_j^a - \alpha_j^b\right)\mu^0$$
$$K_{yy}^j = \left(\alpha_j^a - \alpha_j^b\right)\mu^y$$
$$K_{yz}^j = \left(\alpha_j^a - \alpha_j^b\right)\mu^z \tag{16.16}$$

Now, pick up two arbitrary values for μ^0 and μ^y, with $\mu^y \neq 0$. The last two expressions give $(\alpha_j^a - \alpha_j^b)$ and μ^z; the first gives α_j^b therefore α_j^a.

As expected, a continuum of different models generate the same aggregate demand. Moreover, these differences are welfare relevant, in the sense that the individual welfare gains of a given reform (say, a change in prices and incomes) will be evaluated differently by different models. In practice, the collective indirect utilities recovered above are not invariant across the various structural models compatible with a given aggregate demand.

A unitary version of the model obtains when the Pareto weights are constant: $\mu^y = \mu^z = 0$. Then $K_{yz}^j = 0$ for all j (because distribution factors cannot matter), and $K_{yy}^j = 0$ for all j (demand must be linear in y, as a quadratic term would violate Slutsky). We are left with $K_y^j = \alpha_j^b + (\alpha_j^a - \alpha_j^b)\mu^0$, and it is obviously impossible to identify independently α_j^a, α_j^b, and μ^0; as expected, the unitary framework is not identifiable.

16.5.1.5.3 Identification Under Exclusion

We now show that in the nonunitary version of the collective framework, an exclusion assumption per member is sufficient to exactly recover all of the (welfare-relevant) coefficients. Assume that member a does not consume commodity 1, and member b does not consume commodity 2; that is, $\alpha_1^a = \alpha_2^b = 0$. Then equations (16.15) give

$$\alpha_1^b\left(1-\mu^0\right) = K_y^1, \quad -\alpha_1^b \mu^y = K_{yy}^1, \quad -\alpha_1^b \mu^z = K_{yz}^1$$

and

$$\alpha_2^a \mu^0 = K_y^2, \quad \alpha_2^a \mu^y = K_{yy}^2, \quad \alpha_2^a \mu^z = K_{yz}^2.$$

Combining the first two equations of each block and assuming $\mu^y \neq 0$, we get

$$\frac{1-\mu^0}{\mu^y} = -\frac{K_y^1}{K_{yy}^1} \quad \text{and} \quad \frac{\mu^0}{\mu^y} = \frac{K_y^2}{K_{yy}^2};$$

therefore, assuming $K_y^2 K_{yy}^1 - K_y^1 K_{yy}^2 \neq 0$,

$$\frac{1-\mu^0}{\mu^0} = -\frac{K_y^1 K_{yy}^2}{K_y^2 K_{yy}^1} \quad \text{and} \quad \mu^0 = \frac{K_y^2 K_{yy}^1}{K_y^2 K_{yy}^1 - K_y^1 K_{yy}^2}$$

It follows that

$$\mu^y = \frac{K_{yy}^2}{K_y^2}\mu^0 = \frac{K_{yy}^2 K_{yy}^1}{K_y^2 K_{yy}^1 - K_y^1 K_{yy}^2}$$

and all other coefficients can be computed as above. It follows that the collective indirect utility of each member can be exactly recovered, which allows for unambiguous welfare statements. As mentioned above, identifiability is only generic in the sense that it requires $K_y^2 K_{yy}^1 - K_y^1 K_{yy}^2 \neq 0$. Clearly, the set of parameters values violating this condition is of zero measure. Also, identifiability requires $\mu^y \neq 0$; in particular, *it does not hold true* in the unitary version, in which $\mu^y = \mu^z = 0$. Indeed, the same exclusion restrictions property as above only allow us to recover $\alpha_1^b(1-\mu^0) = K_y^1$ and $\alpha_2^a\ \mu^0 = K_y^2$; this is not sufficient to identify μ^0, let alone the α_j^i for $j \geq 3$. This confirms that the unitary version of the model is not identified even under the exclusivity assumptions that guarantee generic identifiability in the general version.

Finally, one can readily check the previous claim that the MMWIs are not identified. Indeed, the MMWI m^s of s is defined by

$$v^s(\pi, m^s) = \log\left(m^s - \sum_k \pi_k \gamma_k^s\right) - \sum_i \alpha_i^s \log \pi_i = W^s(\pi, y, z),$$

where

$$v^s(\pi, P, y) = \log\left(y - \sum_k \pi_k \gamma_k^s\right) - \sum_{i=1}^n \alpha_i^s \log \pi_i$$

and

$$W^s(\pi, z) = \log\left(y - \sum_{i,k} \pi_i \gamma_i^k\right) + \log\left(\mu^0 + \mu^y y + \mu^z z\right) - \sum_i \alpha_i^s \log \pi_i.$$

This gives

$$m^s(\pi, y, z) = \left(\mu^0 + \mu^y y + \mu^z z\right)\left(y - \sum_i \pi_i\left(\sum_k \gamma_i^k\right)\right) + \sum_i \pi_i \gamma_i^s.$$

For any private commodity i, the sums $\sum_k \gamma_i^k$ are identified, but the individual γ_i^s are not; therefore, m^s is identified up to an additive function of the prices of private, nonexclusive goods.

16.5.2 Comparing Different Family Sizes

A second approach enlarges the set of usable information by allowing comparisons between families of different compositions. A first idea is to assume some relationship between individual preferences when married and single. In that sense, the "pure" approach just described relies on an extreme version. This is because it does not postulate *any* link between utilities when married and single; hence, knowledge of an individual's preferences when single brings no information about her tastes within the household. At the other extreme, some models assume that preferences are unaffected by marital status, at least ordinally. This means that if $u_S^a(Q, q^a)$ denotes a's utility when single, then her utility when married takes the form

$$u^a(Q,q^a) = F\left(u_S^a(Q,q^a)\right),$$

where F is an increasing transform. Thus, marriage can directly affect a person's utility level, but not the person's marginal rates of substitution between various commodities. Note that if we assume preferences are unaffected by marital status, then the MMWI defined above has a natural interpretation; namely, it is the level of income that would be needed by the individual, *if single*, to reach the same utility level as what she currently gets within marriage. It must however be stressed that the assumption of constant preferences across marital status is not needed for the *definition* of the index, but only for this particular interpretation.

Various, intermediate approaches can be found in the literature. One, mostly used in a labor supply context, only assumes that some preference parameters are common to singles and households, and can therefore be estimated separately on a sample of singles. In general, this is sufficient to identify (or calibrate) the remaining parameters (relevant for marriage-specific preferences and the Pareto weights) on observed labor supplies of men and women in a sample of couples. This approach has been adopted in a series of papers recently published in the *Review of Economics of the Household* (Bargain et al., 2006; Beninger et al., 2006; Myck et al., 2006; Vermeulen et al., 2006). For instance, consider a model of labor supply in a couple in which the utility of agent a takes the form

$$u^a\left(q^a, L^a, L^b\right) = \alpha^a \ln\left(q^a - \overline{q}^a\right) + \beta^a \ln\left(L^a - \overline{L}^a\right) + \gamma^a \ln\left(L^a - \overline{L}^a\right) \ln\left(L^b - \overline{L}^b\right),$$

where L denotes leisure; note that this form is more general than the ones considered above, because it allows for (positive) externalities of leisure within the couple.[14] The α and β parameters are assumed to be independent of marital status and are therefore identified from a sample of singles; the γs and the Pareto weights are then calibrated from data on households.

An intermediate approach, that relies on the notion of domestic production, has recently been proposed by Browning et al. (2013). It posits that agents, when they get married, keep the same preferences but can access a different (and generally more productive) technology. That is, while the basic rates of substitution between *consumed* commodities remains unaffected by marriage (or cohabitation), the relationship between purchases and consumptions is not; therefore, the structure of demand, including for exclusive commodities (consumed only by one member) is different from what it would be for singles. More generally, one can, following Dunbar et al. (2013), only assume that preferences are unaffected by family composition (e.g., that parents' preferences regarding their own consumption does not depend on the number of children). These approaches are described in the next section.

[14] Equivalently, this approach considers both leisures as public goods within the household.

16.5.3 Identifying from Market Equilibrium

Lastly, a series of recent contributions are aimed at taking to data the aforementioned equilibrium approaches. The basic, theoretical intuition is quite straightforward: the equilibrium conditions on the marriage market (with or without search frictions, but with intrahousehold transfers) either constrain or exactly pin down intrahousehold allocations. Several papers propose an empirical implementation of this idea. A first set of works only consider matching patterns; the marriage market equilibrium is then exclusively characterized by a matrix of intermarriages between various categories, which can be defined by age, education, income, or any combination of these. On the matching front, following the initial contribution by Choo and Siow (2006), Chiappori et al. (2011) have shown how a structural, parametric model can be (over)identified from such patterns, under the assumption that, while the surplus generated by marriage may (and does) vary over time, its supermodularity (which drives the extent of assortative matching in the population) is constant.[15] According to their estimate, while the gains from marriage have globally decreased over the last decades, the decline has been much smaller for educated couples. Moreover, the share of household resources received has increased for college-educated wives, resulting in a strong increase in their "marital college premium" (defined as the additional gain provided by a university education on the marriage market). This is compatible with the theoretical analysis of Chiappori et al. (2009), who argued that the asymmetry between male and female marital college premiums could explain (at least in part) the higher demand for university training by women. Alternatively, Jacquemet and Robin (2011) and Goussé (2013) analyze marital patterns from a search perspective.

A clear limitation of these approaches is that the sole observation of marital patterns conveys only limited information on the form of the marital surplus (therefore on distribution). For instance, knowing that matching is assortative tells us only that the surplus is supermodular. The previous approaches, therefore, must rely on strong and largely untestable assumptions on the precise form of the heterogeneity distribution across couples. Adding information on the total surplus would greatly enhance the identification power of these models. But such information is precisely what collective models can provide based on observed behavior. The intuition, here, is that the observation of, say, labor supply patterns of married couples (which reflects intrahousehold transfers), together with that of marital patterns, should allow us to fully identify a general matching model in a very robust way. This line of research is pursued, in a series of paper, by Chiappori et al. (2014).

[15] For a general presentation of the econometrics of matching models, see the survey by Chiappori and Salanié (2014).

16.6. EMPIRICAL FINDINGS

In this section we review some empirical work based on the collective model and emphasizing the identification of the sharing rule.

The first-generation models used information on private and assignable goods such as consumption of clothing or individual leisure to identify the sharing rule up to a constant. These models adopt mainly two approaches for identification. The first approach refers to what we called "pure" identification, that is it recovers the derivatives of the sharing rule with no further information than observed consumption bundles of the household. As discussed above, while some identifying conditions can be relaxed by using distribution factors, these models cannot identify separately the level of sharing (how much goes to each household member) from preferences. There exists a continuum of allocations of resources, each associated to a utility function for each household member, that fit the data equally well; across these allocations, income inequality within the household is different, although the allocation of welfare to each member remains the same.

To identify the way overall resources are allocated and thus measure inequality, one needs more information, either in terms of identifying assumptions on the behavior of the sharing rule (such as nonnegativity conditions discussed earlier) or assumptions on preferences. One possibility is to compare the behavior of married and single individuals by making assumptions about the way preferences change with marriage. Other approaches involve specific restrictions on preferences. We show how some of these approaches have been used in the literature. Finally, we also consider the information content of revealed preference restrictions. These extend the revealed preference arguments for individual choice to the case of collective households. Clearly this is a much more complicated setup than standard revealed preference restrictions for individuals or for unitary households because the aggregate household does not necessarily behave like a rational single agent. We discuss what can be learned from revealed preference in this context.

However, the issue of identification of the sharing rule is deeper than what is suggested by the use of the restrictions above and has to do with the way people make agreements at the point of marriage and the level of commitment associated with these agreements. In other words, fundamentally the sharing rule is identified from behavior without having to impose possibly *ad hoc* restrictions. Identification requires extending the model to include marital decisions in an equilibrium context. Indeed a marriage market equilibrium will define the sharing rule, and conditions in the marriage market can allow us to identify it. This effectively introduces dynamics, which then allows one to delve deeper into the extent of commitment and what this means about within-household inequality. Characterizing the theoretical and empirical power of using marriage-market data to better understand intrahousehold allocations is a relatively new and active area of research, particularly when limited commitment is allowed for.

Before we discuss the empirical literature we need to introduce a distinction between the concept of identifiably of preferences and the sharing rule on the one hand and econometric identification on the other. The identifiability results discussed earlier in the chapter relate to our ability to recover individual preferences and the sharing rule given *we know* the household level demand functions exactly. Empirical analysis is concerned with estimating these household demands from empirical data to be able to then recover the sharing rule. This issue brings forth all the standard econometric concerns, such as the role of unobserved heterogeneity, the endogeneity of wages, prices and income, corner solutions (particularly in labor supply), and the like. One of the hardest issues concerns the way that unobserved heterogeneity enters household demands, particularly if such unobservables are correlated with observables. The specific issue arises from the fact that, in general, unobserved heterogeneity in preferences will imply unobservables in the sharing rule. In most specifications this will mean that unobservables are not separable from observables, which has implications for econometric identification. For example, Blundell et al. (2007) used linearity to bypass the difficulties implied by unobserved heterogeneity in preferences. Here we are not offering any general solution to the problem, but we need to point out that, before we even consider identification of the sharing rule, an empirical approach would have to solve the standard econometric identification issues, which in this context may be severe.[16]

16.6.1 "Pure" Identification of the Sharing Rule

In this first generation of collective models we can point to three main empirical studies. The first is by Browning et al. (1994); the second is by Chiappori et al. (2002); and the third is by Blundell et al. (2007). All three share a similar approach to identification: they assume efficiency and an assignable good. However, the details of the empirical approach differ.

In Browning et al. (1994), the authors use a sample of couples drawn from the Canadian FAMEX and estimate a model for the demand of men's and women's clothing, and identify the sharing rule, up to a constant. Identification relies on two assumptions: first, clothing is an assignable good, which effectively means that we can observe male and female clothing and that only the person using the clothing derives any utility from it. In other words, clothing does not include a public good element. Second, they assume that the distribution of a partner's income does not affect preferences, but they may enter the sharing rule, reflecting bargaining positions. Given these assumptions, they identify a sharing rule as a function of the age difference of the partners, total household expenditure (thus allowing wealth effects in the way resources are shared), and most importantly, the share of income attributable to the female partner. It turns out that the way resources are distributed between couples is not very sensitive to the proportion of income for which they are accountable. For example, going from a share of income of 25–75% raises

[16] For recent attempts in this direction (including a discussion of the specific difficulties it raises), see for instance Lewbel and Pendakur (2013) and Chiappori and Kim (2013).

the share of household expenditure by a significant but small 2.3%. The age difference and the level of expenditure also matter with relatively older individuals gaining more and wealthier households allocating more to the wife.

The Browning et al. (1994) paper shows the potential of the approach and the richness of the empirical results that can be obtained by judicious use of information reflecting bargaining power of households. However, the main determinant of female bargaining power in their model is the relative magnitude of female income. A higher share of income may reflect her relative skills or, alternatively, it may reflect her decision to forgo leisure and work more; in other words, this distribution factor is indeed endogenous. In principle, this fact does not harm identification provided that labor supply is separable from consumption: Controlling for total expenditures, individual consumption should then be independent of labor supply (therefore of labor income). However, separability is a strong assumption, that has been empirically criticized. The next two papers address exactly this issue by endogenizing labor supply.

The empirical relevance of the discussion above for within-household inequality and allocation of resources is illustrated by Chiappori et al. (2002). They use data from the PSID to estimate a collective labor supply model, where the sharing rule is identified (up to a constant) based on distribution factors. These include the sex ratio (males/females) in the state as measured by the 1990 census as well as by dummy variables indicating the nature of divorce laws.[17] Measuring the sex ratio is, of course, very tricky, both because we need to define the relevant marriage market and because timing may matter. In a full commitment model, for example, the sex ratio at the time of marriage is what is going to matter. However, the sex ratio is unlikely to change vastly over time, and it is probably a good idea to define marriage markets quite broadly rather than too narrowly. The authors also report using the county-level sex ratio with the state level as an instrument, which had little impact on their results. In their model, labor supply is evaluated over one whole year and they consider a sample where both are working. So the relevant group are individuals with sufficient attachment to the labor market to want to work at least some part of the year. In their model, the sharing rule is allowed to be a function of the wages of both partners, nonlabor income, and the distribution factors. Allowing both wages to enter is important: It has been empirically observed that both wages matter when estimating family labor supply (see, e.g., Blundell and Walker, 1986) a fact that in a unitary context has been interpreted as nonseparability in household preferences between male and female leisure. Here this nonseparability is interpreted as being driven by the impact of the sharing rule on individual labor supply in a collective setting. The

[17] The intuition underlying the CFL paper—that a relative scarcity of women and/or more favorable divorce laws should improve the wife's Pareto weight—can be supported by an explicit matching model, with some nuances (e.g., changes in divorce laws affect differently women already married and women getting married after the change in law). On these issues, see Chiappori et al. (2013).

fact that the restrictions from the collective model are not rejected strengthens this interpretation.

The results suggest that marriage and labor market conditions can lead to large differences in the allocation of household resources within a couple. For example, a $1 increase in the female hourly wage rate leads to a transfer at the means of $1600 to the husband, implying that most of the extra income goes to him. However, a $1 increase in his hourly wage rate leads only to $600 transfer to her, implying he keeps the lion's share and does not behave as altruistically (to use the authors' words) as she does. The wage effects are of particular interest because changes in wages and in male–female wage differentials may be a key driver of within-household allocation of resources. Unfortunately, these results are not precisely estimated; we revisit this issue below in our discussion of Blundell et al. (2007) and of Lise and Seitz (2011). Anyhow, a result that stands out in Chiappori et al. (2002) is the impact of the sex ratio. Based on this result, an increase of one percentage point in the sex ratio leads to $2160 transfer to the wife. Noting that the range of the sex ratio in their data is 0.46–0.57, the implication is that from the least favorable to the most favorable marriage market, the transfer can differ by as much as $23,000. Of course, this does not all translate into an increase in consumption because the income effect on labor supply will imply a change in the amount of hours worked, with women who live in marriage markets more favorable for them working less. To obtain a summary of divorce laws, the authors constructed an index ranging from 1 to 4 and indicating the extent to which the divorce laws are favorable to women. Here the effects are particularly strong as well. A one-point increase in the index leads to a transfer of $4310 to the wife, which again is shared between consumption and leisure.

These results are important because they show the extent to which within-household allocation of resources can be sensitive to external conditions affecting the bargaining power of the members of the couple. Noting, for example, that average household income in this data is $48,000, the change that can be induced just because of (admittedly extreme) changes in the sex ratio can amount to almost half of household income.

However, there are a number of empirical issues that were not addressed by the papers already discussed. First, we need to be concerned that the allocation of women across states with different sex ratios is not random with respect to their unobserved preferences for labor supply. This can bias the results if women who live in areas abundant with men tend to have lower labor market attachment. Second, we need to address the issue of precision in the estimation of wage effects, an issue that persists in the Blundell et al. (2007) paper we will discuss below. CLF instrument wages, but the instruments are necessarily quite weak: they rely on a polynomial in age and education as an instrument while (correctly) controlling linearly for education and age in the labor supply function. This leaves higher-order nonlinearity in the profile of wages with respect to age and education to act as excluded instruments that is both difficult to justify theoretically and, at the same time, is not very informative. To solve these empirical issues we will require exogenous events that change wages and the marriage market, something that a newer generation of collective models is now addressing, such as the paper by Attanasio and Lechene (2014) who use experimental variation in female income induced by the Conditional Cash Transfer (PROGRESA) experiment in Mexico to obtain exogenous variation in the relative bargaining position of males and females.

Beyond these difficulties there is one further important issue that the papers we discussed fail to address, namely nonparticipation of women. Given that many women do not work allowing for this possibility and understanding how resources are allocated despite the fact she is not producing in the formal market is a key concern. The Blundell et al. (2007) paper addresses the question of identification and estimation of a collective labor supply model that allows for male and female non participation in the labour market. In addition, it considers the case where the male labor supply decision is discrete (work or not). This restriction is imposed to accommodate the fact that in the United Kingdom (where the data is drawn from) the male hours of work distribution seem discontinuous between 0 and about 35 h per week, with the entire mass of workers concentrated in the full time range. This restriction is not entirely satisfactory, but it may do better justice to the data than assuming hours of work are freely chosen. Thus, the resulting model is one where females make choices both on the intensive and the extensive margin, whereas males choose only on the extensive margin. The authors prove identification of the sharing rule; however, this is only identified (nonparametrically) if at least one of the two household members work. Parametric restrictions provide the rest. In the empirical implementation Blundell et al. (2007) deal with the endogeneity of the wage rate by exploiting the changes in wage inequality across cohorts and education groups. Econometric identification relies on the assumption that whereas the structure of wages changed across education groups and cohorts—a testable assumption—preferences remained unchanged. This implies that changes in work behavior across cohorts and education groups can be attributed to changes in the incentive structure, which is the identification strategy employed by Blundell et al. (1998).

The empirical analysis is conducted on a sample of married couples, observed between 1978 and 2001 in the UK Family Expenditure Survey. The assumptions imposed for identification (over and above efficiency) required only private goods and one assignable good. The assignable good is leisure. Because expenditures on children are not separately observable in the data, and because these are effectively public, the authors exclude all couples with children and then assume that the observed aggregate household consumption reflects the sum of private consumption of each of the two members of the household.

The model implies two different sharing rules depending on whether the husband works or not. They differ by a monotonic transformation, which in their empirical specification acts as an attenuation factor, implying that the husband only gets a fraction of transfers when he is not working. This fraction is 0.71, implying that the derivatives of the sharing rule (as well as the level) are attenuated by that amount when he is not working. Their empirical approach does not use any distribution factors that can be excluded from preferences: the sharing rule depends on male wages, female wages, and unearned income as well as education and age. It turns out that empirically the effect of the female wage on the sharing rule is not well identified. However, the effect of the male wage is precisely estimated. It implies that 88% of an increase in male-market earnings translates into a transfer to the husband if he is working. Inasmuch as there is no intensive margin for the male decision, this translates to a direct impact on his consumption, if he continues to work. If he does not work, the same change in potential earnings translates to a transfer equal to 62% of the potential increase (0.71×0.88). These results imply that when the earnings of a working husband increase, the resulting increase in the consumption of the wife is only small; if potential earnings increase (and he is not working) her consumption declines substantially and he enjoys more of the household resources. Finally, the wife keeps 73% of increases in unearned income. Nevertheless, unearned income is a relatively low fraction of household income.

These results again illustrate that external factors (here the relative wages) can influence the allocation of resources substantially. Unfortunately Blundell et al. (2007) do not provide precise estimates of the effects of female wages, and this hinders an understanding of how the change in the wage structure affected within household allocations. The source of lack of precision is the relatively small sample size where the man does not work. Moreover, allowing both wages and nonlabor income to be endogenous, while important for obtaining consistent estimates that make sense, does affect precision substantially. The paper does demonstrate that one does not need (in principle) distribution factors for identification. However, looking at the empirical problem from the perspective of Chiappori et al. (2002), other environmental factors may be very important in determining allocations and, if they are omitted, they could bias the results. On the other hand, if included they can be allowed to affect preferences as well. Identification does not require they affect the sharing rule alone.

This first generation of models showed the potential of the collective model for identifying allocations of resources within the black box of the household. However there are key issues that had not been dealt with. First, taxes and welfare were ignored. At one level this is an empirical specification issue because ignoring taxes can bias the estimates of the preference parameters. But at a more fundamental level by not taking into account the tax and welfare system we omit one of the most important factors affecting (and sometimes designed to affect) within household allocations. Estimating models that allow for taxes and welfare can then explain how changes in the policy and the market environment can affect the allocation of resources.

The next fundamental issue is that the models described above can only identify the derivatives of the sharing rule, that is, how sharing changes when distribution factors, prices, and unearned income change. This precludes any discussion of the levels of inequality of resources and hence does not allow us to put into perspective the implication of changes that occur over time.

Adding taxes and welfare does not pose any important conceptual problems. In practice it involves allowing for more complex budget sets and solving the model to take into account nonlinear budget sets. An interesting issue is that the welfare and tax system may create a further interdependence in the decisions of husband and wife, over and above that induced by the sharing rule. These issues are considered, for instance, in Donni (2003), who uses a "pure" identification strategy of the type just described, and by Beninger et al. (2006), Myck et al. (2006), and Vermeulen et al. (2006) who use information from singles and couples.

Extending the model to allow identification of the level of the sharing rule does, however, pose conceptual problems. Fundamentally, the sharing rule is identified by the equilibrium in the marriage market. However, barring the use of a complete marriage market equilibrium model one can obtain information on the level of inequality with alternative auxiliary assumptions. One possibility is to use information on singles. This involves restricting the way preferences change with marriage. This is an approach used by Lise and Seitz in an early version of their paper. Another possibility is to assume something about the sharing rule at one point of the wage space, for example, that all resources are shared equally when wages are equal, which is the assumption made in the published version of Lise and Seitz (2011). Finally, one can make assumptions about the functional forms of demand, as in Dunbar et al. (2013). We now look into these empirical studies.

16.6.2 Intrahousehold Inequality Over Time and the Sharing Rule: Lise and Seitz (2011)

Lise and Seitz (2011) use the collective model to first estimate overall consumption inequality (at the individual level) and to then decompose this to between household and within household. The important economic fact is that the distribution of wages in the United Kingdom changed dramatically over the period they consider (1968–2001) both within and between education groups (see Gosling et al., 2000). Moreover, the structure of the marriage market has also changed with increased degrees of marital sorting over time. They thus set up a model of male and female labor supply with many (but discrete) choices of hours worked for both members of the household. Hours can take values from 0 to 65 in 5-h intervals. In many ways their empirical framework is similar to that of Blundell et al. (2007): They use couples with no children drawn from the UK Family Expenditure Survey over many years. However, they depart in a number of important ways. First, they allow for taxes and account for the impact of joint taxation over the period that this was in effect in the UK (up to 1989). Also they allow for a richer choice set for the male and they impose further structure so as to identify the level of the sharing rule as well as its derivatives. Finally they account for public goods when they define consumption, although they are taken as separable from private consumption and leisure.

Whereas the logic underlying the identification of the derivatives of the sharing rule is similar to that of Blundell et al. (2007), identification of the location (level) of the sharing rule empirically is based on the identifying assumption that when individuals have the same potential earnings they share resources equally. In earlier versions of the paper it was instead assumed that preferences of married and single individuals are identical; both these assumptions can identify the model. The point at which one pins down the sharing rule is welfare irrelevant, because the preference specification adapts to leave welfare unchanged when the location of sharing is fixed. In principle, just normalizing the location parameter will not cause any bias but will, of course, lead to a specific level of inequality. On the other hand, using information from singles has the advantage that it uses a restriction grounded in some explicit assumption on preferences (marriage does not affect marginal utilities) but, if wrong, will bias all results.

Over the period considered in the paper (1968–2001), earnings inequality increased rapidly. There has been a steady increase in both the potential earnings' and actual earnings' share of women relative to men and a decline in male employment while female employment increased at the start of the period later remaining constant. Consumption inequality increased rapidly in the period between 1980 and 1990, but was basically stable the rest of the time. When Lise and Seitz interpret these results under the prism of their collective model, they uncover some interesting facts: while between-household inequality of consumption increases, within-household inequality of consumption declines to such an extent that the overall inequality of consumption remains more or less the same over time. When they consider a different measure of resources, namely full consumption, which includes the value of leisure enjoyed by each member, they find similar but less stark results. First, between-household inequality still increases, but much less dramatically because the decline in consumption for those households who have workless members is compensated by the value of leisure. Second again they find that within-household inequality declines as before, but much less. Obviously none of these consumption measures is ideal and a money-metric measure of welfare may be better. However, these results illustrate exactly the potential importance of finding credible ways to understand inequality (and poverty) within households. This is more so given that who marries whom is endogenous and in part drives the way that within-household inequality is determined and has implications for between-household inequality is determined.

16.6.3 Intrahousehold Inequality and Children

While intrahousehold inequality may be of general interest because it tells us about allocation of resources within a household and can reveal hidden poverty and inequality, the whole issue acquires special importance when it comes to allocations of consumption to children. Thus is because child consumption and more generally investments in children have long-term implications for the intergenerational transmission of poverty. Yet little or no empirical work had been done to understand how resources are allocated to children and the extent to which reallocations of income from the male spouse to the female can affect the shares directed to children. A theoretical framework for the analysis of this question has been developed by Blundell et al. (2005). In a recent important paper, Dunbar et al. (2013) address this issue empirically using data from Malawi. In their model, each child is represented as having his or her own utility function. This creates a very special difficulty regarding the assumption, used for identification in studies such as Browning et al. (2013), namely that preferences of singles and married individuals are the same. Here, such a strategy is no longer available because children are never seen living as singles. Moreover, in data from Malawi that the authors use, there is not enough price variation—another requirement of the Browning et al. approach. Thus, identification is obtained by making assumptions on the structure and shape of the Engel curves.

The identification strategy first requires either one assignable private good or one exclusive good per person. Remember that an exclusive good is exclusively consumed by one household member type (for example, child clothing is consumed only by children), whereas an assignable good is such that each member's consumption of this good is observable.[18] Of course, there can be many other purely private goods (such as food) for which we do not observe the amounts of individual consumption: This fact does not hamper identification.

The assignability assumption is not sufficient to identify the share of resources of each household member; additional assumptions are therefore needed. Dunbar et al. (2013) assume, first, that resource shares are invariant to total expenditure. In addition, they make two alternative assumptions on preferences: either the demand for goods is similar across household types (i.e., households with one, two, or more children) or they are similar across types of goods within a household type. An extreme form of the assumption is that preferences do not vary across types of household; since shadow prices vary across households because of the partially public nature of goods, this extreme assumption is essentially equivalent to assuming that the assignable good used for identification is irresponsive to prices. Another extreme form of this assumption is that preferences over the assignable good are identical across different household member types (male, female, and children). However, Dunbar et al. (2013) show that identification only requires that some aspect of the demand functions be the same either across household member types

[18] It should be stressed that a good is private when its consumption does not alter the *preferences* of other household members over goods consumption. As the authors put it, smoking by one household member may annoy the others, but it can still be taken as private if it does not in itself alter their consumption of goods.

or across household types. Thus, in one case, they assume that all household members share the same shape of Engel curves for the assignable good. In another case they assume that preferences are the same across types of household (number of children), conditional on a deflator of income. This deflator reflects the different shadow prices that different-sized households face and is the way that preferences for the assignable good are allowed to vary across types. The key point is that the authors need to define similarity so that identification is delivered without sacrificing theoretical consistency (integrability) of the demand functions.

Dunbar et al. (2013) estimate their model on data from Malawi, probably in one of the first such studies with development data. In a sense their framework is very well adapted to this context because wages and/or prices, which are at the heart of some other identification strategies and are not observed in that case. Their approach relies on measuring expenditures and having an assignable good for which they use clothing and footwear. The results they obtain are both astounding and an excellent illustration of the importance of looking within the household. They find that the male obtains about 45–50% of household resources. His share seems to be insensitive to the number of children present. The mother's share declines with the second child, but then remains more or less constant, with the consumption share of children declining.

Even more pertinent are the implied poverty rates. Male poverty rates are at their highest in one-child households and seem to decline in households with more children. However, the important result concerns poverty rates for women and children: Compared to the male poverty rate of around 69%, there are 79% poor women and 95% poor children in one-child households. In larger households, the male poverty rate is about 55% whereas the female poverty rate is 89% and nearly all children are poor. Hence their approach not only offers a more complete picture of poverty but reveals the extent of child poverty, which is crucial to development. Without such an approach, child poverty would not be apparent to the extent that it is in reality. Although the authors did not focus on gender differences between children, which may be another important dimension, this line of research can easily be extended in that direction; it offers an obvious mechanism for trying to understand how resources are allocated by gender.

A potential limitation of this approach is the fixed nature of the sharing rule. While the authors spend a lot of time explaining the upsides of not relying on distribution factors (essentially, they avoid having to take a position on whether they affect preferences or not), the absence of an underlying model of what the resource share should depend on and how it can be affected by exogenous driving forces may in some cases be problematic. In models where the sharing rule is allowed to depend on wages or institutional features we have some understanding of how policy can be used to target individuals. In the Dunbar et al. model this aspect is missing. However, this is not an integral part of the approach, and richer models can be identified.

16.6.4 Revealed Preference Restrictions and the Identification of the Sharing Rule

The approach to the identification of the sharing rule has exploited the structure of the demand functions and the way that income affects observed outcomes when the collective model is true. This leads to a set of differential equations that when solved provide the derivatives of the sharing rule. As already discussed, this is not sufficient for identifying the level of the sharing rule.

A different approach is that of revealed preference. In the context of the single-agent utility maximization model the axioms of revealed preference allow one to test nonparametrically whether a particular set of choices can be rationalized by utility maximization and if they can, to bound the underlying demand functions. Such an approach has been developed and implemented for the unitary model by Blundell et al. (2003, 2008) and is based on the original work of Afriat (1973) and Varian (1982). In the collective framework the aggregate household demands will in general violate the revealed preference restrictions corresponding to the unitary model simply because as the budget constraint changes (wages, prices, incomes, etc.) individuals make different choices and the Pareto weights change. This insight was developed by Browning and Chiappori (1998), who showed that the aggregate household demands have to possess a Slutsky matrix that can be decomposed into a symmetric matrix plus a matrix with rank equal to the number of decision makers (whose demands are aggregated) minus 1. The fact that the pattern of choices is restricted implies that there should also be revealed preference type restrictions, as noticed by Chiappori (1988), who provides an early example in a labor supply context. Indeed these restrictions have been fully developed by Cherchye et al. (2007). In a further development Cherchye et al. (2012) show how the revealed preference restrictions can be used to bound the sharing rule without imposing any restrictions other than Pareto efficiency of intrahousehold allocations. The main result is based on the following principle: Suppose that a set of observed demands are collectively rationalizable in the sense that the observed choices are consistent with the existence of admissible individual demand functions. Then it has to be that any alternative choices that could lead to a Pareto improvement within the household should be infeasible at current market prices and for any allocation of income within the household such that each person receives a nonnegative share. More specifically, consider the set of demands that individual 1 reveals, based on all possible admissible demand functions for that person. They must cost more than person 1's share of total household income; similarly for person 2. The least-costly bundle that would lead to a Pareto improvement provides the upper bound for a person's share. Adding up the shares to total income and the assumption that the shares cannot be negative determines the lower bound. The difficulty in implementing this principle is the fact that we need to search over all possible admissible individual demand functions.

This principle turns out to generate nontrivial upper and lower bounds for the sharing rule. Importantly, no restriction is needed for such bounds other than Pareto optimality: All or some goods may be either private, in part public, and in part private or completely private. Moreover, we do not need to specify which goods (if any) are purely private, but if such information were to be available it can be used to tighten the bounds.

Cherchye et al. apply their approach to the PSID from 1999, when expenditures on individual consumption goods became available, until 2009. The sample consists of childless couples where both are working. Utility depends on leisure, food and other goods, which include health and transportation. Leisure is assumed to be assignable, but no assumption is made on the other goods. This is important because in this case, at least in general, neither the level nor the derivatives of the sharing rule are point identified.

To implement their approach they start by estimating three different versions of an aggregate household demand system: a nonparametric system, the QUAIDS demand system (Banks et al., 1997) and a QUAIDS demand system where the substitution matrix is restricted to be symmetric plus rank one, which imposes that the demands are consistent with the collective model. Given this demand system they apply their algorithm to bound the sharing rule for different values of the full household income, wages, and prices. Their empirical results are remarkable. First, the bounds are very narrow with the nonparametric demand system implying 12% median difference between upper and lower bounds and the fully restrictive demand system implying only 3%. Going from the nonparametric demand system to the unrestricted QUAIDS system, the tightening is due to imposing the parametric restrictions that may or may not be valid—the authors provide no evidence on that matter. However, assuming the parametric restrictions are valid, the further step of going from QUAIDS to restricted QUAIDS is just imposing restrictions that are implied by the problem and hence serve only to make the bounds sharp(er). Thus, when Pareto efficiency is imposed, the median difference between the upper and lower bound tightens from about 9% to 3%, a substantial improvement. It would have been useful to use a shape-constrained nonparametric demand system (see Blundell et al., 2012) avoiding the parametric restrictions, but using the Pareto constraints as implied by the model.

Using their bounds they establish that the female share is a normal good, that is, as full-household income grows so does the female share; interestingly, this finding confirms results previously derived in different contexts. Moreover they show that in percentage terms the average female share is very closely bounded around 50%, although there is substantial heterogeneity around that point. However, it is impressive how tightly bounded the sharing rule is throughout the distribution. In interpreting this result, one needs to be careful because it is full income that is being shared equally. This measure of income includes both leisure and consumption. Thus the share of a woman with a high wage who does not work will include her leisure and her consumption; hence a 50% share may in certain cases hide very unequal levels of consumption of all other goods.

In the final part of the analysis, the authors use their estimates to carry out a poverty analysis. The idea here is similar to that in Dunbar et al. (2013) described earlier: They compare poverty rates implied by household-level income and those implied by individual allocations. The household poverty line is 60% of median household income whereas the individual poverty line is set at half this amount. This, of course, is an income-based and not a welfare-based measure and ignores any household economies of scale. This point notwithstanding, the individual rates are higher: while household poverty is 11%, individual poverty is bounded between 16% and 21%, the lower bound being above the household number. Interestingly the bounds do not differ by gender by any substantive amount.

The Cherchye et al. study breaks new ground and shows the power of the collective approach. Specifically it reinforces the identifiability results substantially by showing not only that the levels of the sharing rule can be identified, but more importantly in our view, that the entire sharing rule can also be bounded without much more than within-household Pareto efficiency. Nevertheless, there is still a long and important agenda in this research. First, empirically we need to understand better how to deal with heterogeneity in preferences within such a nonparametric framework as well as with endogeneity of prices and wages. The entire analysis of Cherchye et al. is based on the assumption that wages and prices are exogenous. This is internally consistent with the absence of heterogeneity and shocks, but is broadly unsatisfactory. For example, there is a vast labor supply literature dealing with endogenous wage rates. Moreover, prices of goods may not be exogenous if there are aggregate shocks to the demand functions. While these seem to be side issues as far as the central identifiability of the collective model is concerned, they are important for the ultimate empirical credibility of the approach.

16.7. CONCLUSION

Understanding intrahousehold inequality and, more broadly, intrahousehold allocations is crucial for understanding the effects of policy and for targeting programs designed to alleviate poverty. The implications are far reaching and they span simple questions of who will benefit from certain programs to deeper questions about child poverty and even child development. It is now well understood that treating households as an individual unit does not just provide an incomplete picture of standards of living but can be seriously misleading when we try and understand behavior and its reactions to the environment. In our review we have discussed both the questions underlying the notion of intrahousehold inequality as well as the extent of our ability to identify what goes on in the household from typically observed data. In this context we have argued that it is important to be able to observe variables that shift the bargaining power of spouses without affecting preferences as well as other approaches to peeking inside the household black box. It is evident from this discussion that better data would be important; and nothing is more important than detailed consumption and time use data. A renewed emphasis on such data is called for, given the importance of the issues at hand. A better understanding of what may constitute distribution factors and indeed experimental evidence would be an important way to support research into intrahousehold allocations.

However, beyond the above, research is now advancing into the dynamics of intrahousehold allocations and being linked to marriage markets. It is now becoming clear how the conditions at the time of marriage can affect intrahousehold allocations. Indeed, under full commitment, current distribution factors may have little to do with current allocations. On the other hand, full commitment is a very strong and some may argue an implausible assumption. Thus, research is also advancing in understanding how allocations are determined when commitment is limited. In such limited commitment environments changes in the institutional framework, such as the structure of the welfare system or divorce laws, may have important implications for intrahousehold inequality as well as for the formation and dissolution of marriages. We thus are acquiring a rich theoretical and empirical framework that will allow us to better understand how individual welfare is determined within the context of the family. Important contributions in understanding the dynamics of intrahousehold allocations and of household formation include papers by Mazzocco (2007) and Voena (2013). We are convinced that this is a crucial direction for future research.

ACKNOWLEDGMENTS

P.-A. C. gratefully acknowledges financial support from the NSF (grant 1124277). C. M. is grateful for financial support by the Cowles foundation and the Institution for Social and Policy Studies at Yale. Moreover, we thank Tony Atkinson, Francois Bourguignon, and Marc Fleurbaey for useful comments on a previous version. The usual disclaimer applies.

REFERENCES

Afriat, S.N., 1973. On a system of inequalities in demand analysis: an extension of the classical method. Int. Econ. Rev. 14, 460–472.

Attanasio, O., Lechene, V., 2014. Efficient responses to targeted cash transfers. J. Polit. Econ. 122 (1), 178–222.

Banks, J., Blundell, R.W., Lewbel, A., 1997. Quadratic Engel curves and consumer demand. Rev. Econ. Stat. 79 (4), 527–539.

Bargain, O., Beblo, M., Beninger, D., Blundell, R., Carrasco, R., Chiuri, M.-C., Laisney, F., Lechene, V., Moreau, N., Myck, M., Ruiz-Castillo, J., Vermeulen, F., 2006. Does the representation of household behavior matter for welfare analysis of tax-benefit policies? An introduction. Rev. Econ. Househ. 4, 99–111.

Basu, K., 2006. Gender and say: a model of household behavior with endogenously-determined balance of power. Econ. J. 116, 558–580.

Becker, G.S., 1974. A theory of social interactions. J. Polit. Econ. 82 (6), 1063–1093.

Beninger, D., Bargain, O., Beblo, M., Blundell, R., Carrasco, R., Chiuri, M.-C., Laisney, F., Lechene, V., Longobardi, E., Moreau, N., Myck, M., Ruiz-Castillo, J., Vermeulen, F., 2006. Evaluating the move to a linear tax system in Germany and other European countries. Rev. Econ. Househ. 4, 159–180.

Blundell, R.W., Walker, I., 1986. A life-cycle consistent empirical model of family labour supply using cross-section data. Rev. Econ. Stud. 53 (4, Econometrics Special Issue), 539–558.

Blundell, R.W., Duncan, A., Meghir, C., 1998. Estimating labor supply responses using tax reforms. Econometrica 66 (4), 827–861.

Blundell, R.W., Browning, M., Crawford, I.A., 2003. Nonparametric Engel curves and revealed preference. Econometrica 71 (1), 205–240.

Blundell, R.W., Chiappori, P.A., Meghir, C., 2005. Collective labor supply with children. J. Polit. Econ. 113 (6), 1277–1306.

Blundell, R.W., Chiappori, P.-A., Magnac, T., Meghir, C., 2007. Collective labour supply: heterogeneity and non-participation. Rev. Econ. Stud. 74 (2), 417–445.
Blundell, R.W., Browning, M., Crawford, I.A., 2008. Best nonparametric bounds on demand responses. Econometrica 76 (6), 1227–1262.
Blundell, R., Horowitz, J.L., Parey, M., 2012. Measuring the price responsiveness of gasoline demand: economic shape restrictions and nonparametric demand estimation. Quant. Econ. 3 (1), 29–51.
Browning, M., Chiappori, P.A., 1998. Efficient intra-household allocations: a general characterization and empirical tests. Econometrica 66 (6), 1241–1278.
Browning, M., Chiappori, P.A., Lechene, V., 2010. Distributional effects in household models: separate spheres and income pooling. Econ. J. 120 (545), 786–799.
Browning, M., Chiappori, P.A., Lewbel, A., 2013. Estimating consumption economies of scale, adult equivalence scales, and household bargaining power. Rev. Econ. Stud. 80 (4), 1267–1303.
Browning, M., Chiappori, P.A., Weiss, Y., 2014. Family Economics. Cambridge University Press, New York.
Browning, M., Bourguignon, F. Chiappori, P-A., Lechene, V. 1994. Income and outcomes. J. Polit. Econ. 102 (6), 1067–1096.
Cherchye, L., de Rock, B., Vermeulen, F., 2007. The collective model of household consumption: a nonparametric characterization. Econometrica 75 (2), 553–574.
Cherchye, L., De Rock, B., Vermeulen, F., 2009. Opening the black box of intrahousehold decision making: theory and nonparametric empirical tests of general collective consumption models. J. Polit. Econ. 117 (6), 1074–1104.
Cherchye, L.J.H., de Rock, B., Lewbel, A., Vermeulen, F.M.P., 2012. Sharing Rule Identification for General Collective Consumption Models, Tilburg University, Center for Economic Research, Discussion Paper 2012–041.
Chiappori, P.-A., 1988. Rational household labor supply. Econometrica 56 (1), 63–90.
Chiappori, P.-A., 1991. Nash-bargained household decisions: a rejoinder. Int. Econ. Rev. 32 (3), 761–762.
Chiappori, P.-A., 1992. Collective labor supply and welfare. J. Polit. Econ. 100 (3), 437–467.
Chiappori, P.-A., 2012. Modèles d'appariement en conomie: quelques avances recentes. Texte de la Confrence Jean-Jacques Laffont Revue Economique 63 (2012/3), 437–452.
Chiappori, P.-A., Ekeland, I., 2006. The microeconomics of group behavior: general characteriation. J. Econ. Theor. 130 (1), 1–26.
Chiappori, P.-A., Ekeland, I., 2009a. The micro economics of efficient group behavior: identification. Econometrica 77 (3), 763–799.
Chiappori, P.-A., Ekeland, I., 2009b. The economics and mathematics of aggregation. Foundations and Trends in Microeconomics. Now Publishers, Hanover, USA.
Chiappori, P.-A., Kim, J.-H., 2013. Identifying Heterogeneous Sharing Rules. Mimeo, Columbia University, New York.
Chiappori, P.-A., Meghir, C., 2014. Intrahousehold Welfare, mimeo.
Chiappori, P.-A., Salanié, B., 2014. The econometrics of matching models. J. Econ. Lit., forthcoming.
Chiappori, P.-A., Weiss, Y., 2007. Divorce, remarriage and child support. J. Labor Econ. 25 (1), 37–74.
Chiappori, P-A., Costa-Dias, M., Meghir, C., 2014. Marriage Market, Labor Supply and Education Choice. Mimeo, Columbia University, New York.
Chiappori, P.-A., Donni, O., Komunjer, I., 2012. Learning from a Piece of Pi. Rev. Econ. Stud. 79 (1), 162–195.
Chiappori, P.-A., Fortin, B., Lacroix, G., 2002. Marriage market, divorce legislation, and household labor supply. J. Polit. Econ. 110 (1), 37–72.
Chiappori, P.A., Iyigun, M., Weiss, Y., 2009. Investment in schooling and the marriage market. Am. Econ. Rev. 99 (5), 1689–1717.
Chiappori, P.-A., Salanie, B., Weiss, Y., 2011. Partner Choice and the Marital College Premium. Columbia University Academic Commons.
Chiappori, P.-A., McCann, R., Nesheim, L., 2010. Hedonic price equilibria, stable matching, and optimal transport: equivalence, topology, and uniqueness. Econ. Theor. 42 (2), 317–354.
Chiappori, P.A., Iyigun, M., Lafortune, J., Weiss, Y., 2013. Changing the Rules Midway: The Impact of Granting Alimony Rights on Existing and Newly-Formed Partnerships. Mimeo, Columbia University, New York.
Choo, E., Siow, A., 2006. Who marries whom and why. J. Polit. Econ. 114 (1), 175–201.

Donni, O., 2003. Collective household labor supply: non-participation and income taxation. J. Public Econ. 87, 1179–1198.
Donni, O., 2009. A simple approach to investigate intrahousehold allocation of private and public goods. Rev. Econ. Stat. 91, 617–628.
Dunbar, G.R., Lewbel, A., Pendakur, K., 2013. Children's resources in collective households: identification, estimation, and an application to child poverty in Malawi. Am. Econ. Rev. 103 (1), 438–471 (34).
Edlund, L., Korn, E., 2002. A theory of prostitution. J. Polit. Econ. 110 (1), 181–214.
Fleurbaey, M., Gautier, G., 2009. International comparisons of living standards by equivalent incomes. Scand. J. Econ. 111, 597–624.
Fleurbaey, M., Decancq, K., Schokkaert, E., 2014. Inequality, Income and Well-Being. Elsevier, Amsterdam, this Handbook.
Gersbach, H., Haller, H., 2001. Collective decisions and competitive markets. Rev. Econ. Stud. 68, 347–368.
Gosling, A., Machin, S. Meghir, C., 2000. The changing distribution of male wages in the U.K. Rev. Econ. Stud. 67 (4), 635–666.
Goussé, M., 2013. Marriage Market and Intra-Household Allocation: Evolution of Preferences and Transfers in the UK from 1991 to 2008, mimeo Sciences Po.
Grossbard-Shechtman, S., 1993. On the Economics of Marriage: A Theory of Marriage, Labor, and Divorce. Westview Press, Boulder.
Haddad, L., Kanbur, R., 1990. How serious is the neglect of intrahousehold inequality. Econ. J. 100, 866–881.
Haddad, L., Kanbur, R., 1992. Intrahousehold inequality and the theory of targeting. Eur. Econ. Rev. 36, 372–378.
Hammond, P.J., 1994. Money metric measures of individual and social welfare allowing for environmental externalities. In: Eichhorn, W. (Ed.), Models and Measurement of Welfare and Inequality. Springer, Berlin.
Jacquemet, N., Robin, J.-M., 2011. Marriage with labor supply. CES Working Papers 2011.50.
Lechene, V., Preston, I., 2011. Noncooperative household demand. J. Econ. Theor. 146 (2), 504–527.
Lewbel, A., Pendakur, K., 2013. Unobserved Preference Heterogeneity in Demand Using Generalized Random Coefficients. Boston College, Mimeo.
Lise, J., Seitz, S., 2011. Consumption inequality and intrahousehold allocations. Rev. Econ. Stud. 78 (1), 328–355.
Lundberg, S., Pollak, R.A., 1993. Separate spheres bargaining and the marriage market. J. Polit. Econ. 101 (6), 988–1010.
Manser, M., Brown, M., 1980. Marriage and household decision-making: a bargaining analysis. Int. Econ. Rev. 21 (1980), 31–44.
Mazzocco, M., 2007. Household intertemporal behaviour: a collective characterization and a test of commitment. Rev. Econ. Stud. 74 (3), 857–895.
Mcelroy, M.B., Horney, M.J., 1981. Nash-bargained household decisions: toward a generalization of the theory of demand. Int. Econ. Rev. 22, 333–349.
Myck, M., Bargain, O., Beblo, M., Beninger, D., Blundell, R., Carrasco, R., Chiuri, M.-C., Laisney, F., Lechene, V., Longobardi, E., Moreau, N., Ruiz-Castillo, J., Vermeulen, F., 2006. The working families' tax credit and some European tax reforms in a collective setting. Rev. Econ. Househ. 4, 129–158.
Samuelson, P.A., 1956. Social indifference curves. Q. J. Econ. 70 (1), 1–22.
Ulph, D., 2006. Un modèle non-coopératif de Nash appliqué à l'étude du comportement de consommation du ménage. Actualité économique: revue d'analyse économique 82, 53–86.
Varian, H., 1982. The nonparametric approach to demand analysis. Econometrica 50, 945–974.
Vermeulen, F., Bargain, O., Beblo, M., Beninger, D., Blundell, R., Carrasco, R., Chiuri, M.-C., Laisney, F., Lechene, V., Moreau, N., Myck, M., Ruiz-Castillo, J., 2006. Collective models of labor supply with nonconvex budget sets and nonparticipation: a calibration approach. Rev. Econ. Househ. 4, 113–127.
Voena, A., 2013. Yours, Mine and Ours: Do Divorce Laws Affect the Intertemporal Behavior of Married Couples? mimeo University of Chicago.

CHAPTER 17

Health and Inequality

Owen O'Donnell*,†, Eddy Van Doorslaer‡, Tom Van Ourti*
*Erasmus School of Economics & Tinbergen Institute, Erasmus University Rotterdam, Rotterdam, The Netherlands
†School of Economics and Area Studies, University of Macedonia, Thessaloniki, Greece
‡Erasmus School of Economics, Tinbergen Institute & Institute for Health Policy and Management, Erasmus University Rotterdam, Rotterdam, The Netherlands

Contents

Handbook of Income Distribution, Volume 2B
ISSN 1574-0056, http://dx.doi.org/10.1016/B978-0-444-59429-7.00018-2

Abstract

We examine the relationship between income and health with the purpose of establishing the extent to which the distribution of health in a population contributes to income inequality and is itself a product of that inequality. The evidence supports a substantial impact of ill-health on income, mainly operating through employment, although the magnitude of ill-health's contribution to income inequality is difficult to gauge. Variation in exposure to health risks early in life could be an important mechanism through which health may generate and possibly sustain economic inequality. If material advantage can be exercised within the domain of health, then economic inequality will generate health inequality. In high-income countries, the evidence that income (wealth) does have a causal impact on health in adulthood is weak. But this may simply reflect the difficulty of identifying a relationship that, should it exist, is likely to emerge over a lifetime as poor material living conditions slowly take their toll on health. There is little credible evidence to support the claim that the economic inequality in society threatens the health of all its members or that relative income is a determinant of health.

JEL Classification Codes

D31, I14, J3

Keywords

Income, Wealth, Health, Inequality

17.1. INTRODUCTION

The financially better-off also tend to be in better health. This holds between and within countries, both developed and developing, and it has been evident for a considerable period of time (Hibbs, 1915; Van Doorslaer et al., 1997; Woodbury, 1924). There is an income gradient in mortality, as well as in a variety of measures of morbidity and disability. The income–health relationship is not confined to a health gap between the poor and the rest, however. Health continues to rise with income among the nonpoor.

The strength, ubiquity, and persistence of the positive relationship between income and health make it of considerable interest for those studying distributions of income, health, and well-being. Understanding the mechanisms that generate the income–health nexus can help account for inequality, as well as identify inequity, in each of those distributions. This chapter examines the strength and nature of the relationships between income and health with the purpose of establishing the extent to which the distribution of health in a population contributes to economic inequality and is itself a product of that inequality.

The distribution of health is potentially both a cause and a consequence of the distribution of income. Differences in health can generate differences in income, most obviously by restricting earnings capacity. But health inequality may itself reflect economic inequality if health-enhancing goods, such as medical care and nutritious food, are allocated by price. The potentially bidirectional relationship between health and income is relevant both to the positive explanation of the distribution of income and to its normative evaluation. A full understanding of how income differences across individuals are generated requires identification of the extent to which health constrains income. This positive exercise feeds into the normative one of evaluating the distribution of income because the inequity of income inequality surely depends on its causes. The ethical judgment of income distribution is also contingent on its consequences. If money can buy health, then there may be greater aversion to inequality in the distribution of income than there would be if the rich were merely able to afford smarter clothes and faster cars.

The relationship between income and health is not only of interest to those concerned with the distribution of income. From the public health perspective, attention is drawn to observed increases in health with income, as opposed to the corresponding decrease in income with ill-health. Public health scientists tend to interpret the income gradient in health as a symptom of inequity in the distribution of health (Commission on the Social Determinants of Health, 2008), while economists are inclined to view the gradient as reflecting the operation of the labor market in which the sick and disabled are constrained in their capacity to generate earnings (Deaton, 2002; Smith, 1999, 2004). Resolution of this debate is obviously crucial to the formation of the appropriate policy response to the gradient. If it mainly reflects the impact of ill-health on income, then the proposal to use income redistribution as an instrument of health policy (Commission on the Social Determinants of Health, 2008; Navarro, 2001) would be entirely inappropriate (Deaton, 2002).

The inclusion of this chapter in the Handbook is partly motivated by insights into the explanation and evaluation of income distribution that can be gained through the study of the income–health relationship, but it also reflects a trend away from the more narrow focus on differences in income to the more encompassing analysis of inequality in well-being.

Health and income are typically cited as the most important determinants of well-being and are the most common arguments of multidimensional measures of inequality (see Chapters 2 and 3). For given degrees of inequality in the marginal distributions of income and health, most would consider that inequality in well-being is greater when the poor also tend to be in worse health. Understanding the nature of the relationship between income and health is central to determining the degree of inequality in well-being.

17.1.1 Health to Income

There are multiple mechanisms through which health may affect income distribution, with the labor market obviously being an important one. Differences in productivity deriving from variation in physical and mental capacities related to illness and disability are also potentially important determinants of earnings. Differences in the nature of work and infrastructures may mean that physical disability represents a greater constraint on earnings in low-income settings, and mental health problems are relatively more important in developed countries. Discrimination may further widen any disparity in earnings between the disabled and able-bodied. Institutional constraints on wage flexibility may result in unemployment of less healthy individuals who are less productive or face discrimination. On the supply side, ill-health may shift preferences away from work, and this may be reinforced by reduced financial incentives arising from a lower offer wage and entitlement to disability insurance (DI). The latter will cushion the earnings loss arising from disability and so compress the income distribution, but this will be offset if the financial incentives induce withdrawal from employment at a given degree of disability, which may strengthen the earnings–health relationship, if not the income–health relationship, in high-income countries relative to low-income countries. Beyond its effect on the distribution of personal income, health may impact the distribution of household income through the formation and maintenance of marriage partnerships and spousal earnings given needs for informal care.

The impact of health on income may operate with a very long lag. Poor health in childhood may disrupt schooling. Exposure to health risks *in utero* and illness in infanthood may impair cognitive functioning and reduce the efficiency of education in producing knowledge and skills. Childhood health problems may be persistent, such that less healthy young adults enter the labor market with less human capital and lower prospects for lifetime earnings. Early-life health conditions may impact income not only through human-capital acquisition but also by triggering health problems in adulthood (Barker, 1995) that interfere with work. If exposure to health risks in early life is related to economic circumstances, then childhood health could be partly responsible for the transmission of these circumstances across generations (Currie, 2009). According to this proposition, poorer mothers with less education deliver less healthy babies and raise sicker children who acquire less human capital and suffer persistent health problems, both of which constrain earnings and increase the likelihood of parenting a child with health problems. If this theory is empirically significant, then it would place health policy at the very heart of social policy.

17.1.2 Income and Income Inequality to Health

The distribution of income may have consequences for population health through two broad mechanisms. First, the health of an individual may depend on his or her (parents') level of income. If health is a normal good, then demand for it rises with income, and the relationship should be stronger in countries that rely more on the market to allocate health resources, in particular medical care. Second, some claim that the health of an individual is contingent not only on his own income but also on the economic inequality within the society in which he lives (Wilkinson, 1996; Wilkinson and Pickett, 2010). Aggregate data show a clear negative association between measures of population health and income inequality. One proposed mechanism is that psychosocial stress arising from the stigma attached to low relative incomes is physiologically damaging. But the negative relationship between average health and income dispersion could also arise from decreasing health returns to absolute income (Gravelle, 1998; Rodgers, 1979). We weigh the evidence that not only income but also income inequality has a causal impact on health and so affects the distribution of health in a population.

17.1.3 Scope of the Chapter

The literature on the socioeconomic determinants of health is immense and comes from epidemiology, sociology, demography, and psychology, as well as economics. We confine attention to the relationship between income and health, which has been the focus of the economics discipline. Our goal is to establish what is known about the relationship from empirical analyses, and we do not cover the normative literature on health inequality. The evidence we assemble is relevant to the ethical judgment of distributions of income, health, and well-being, but we do not discuss how such normative evaluations might be conducted. Interested readers can consult the excellent discussion of some of the normative issues by Fleurbaey and Schokkaert (2011), as well as Chapters 2 and 4. Relatedly, we do not cover the burgeoning literature on the measurement of income-related health inequality (Erreygers and Van Ourti, 2011; Van Doorslaer and Van Ourti, 2011). Inequality in both income and health could also be analyzed using measures of multidimensional inequality, which are discussed in Chapter 3.

Population health is a standard covariate in empirical growth models, and its contribution to growth has been the focus of a substantial literature aiming to estimate the economic returns from health investments (Barro, 2003, 2013; Commission on Macroeconomics and Health, 2001). We do not cover this literature on the relationship between average income and health because it says nothing about the distribution of each variable across individuals. We do cover evidence on the impact of individual health on income and of income on health, looking at low-income, as well as high-income, countries. But the balance is tilted toward a focus on the latter. Comprehensively covering the very large literature on the impact of health (and nutritional status) on earnings in low-income settings (Strauss and Thomas, 1998) would be too unwieldy. We refer to this literature mainly to establish whether the income–health relationships observed in this setting differ from those in high-income economies that we consider in more detail.

Although we have referred until now to the relationship between income and health, our scope is a little broader. We also consider the relationships between wealth and health. Wealth is an economic outcome of intrinsic interest and is arguably a more appropriate indicator of the economic status of older individuals who provide much of the action in terms of variation in health. The health–wealth effect is likely to differ from the health–income effect. Health may affect income largely through labor market returns. This will feed through to the distribution of wealth, but, in addition, ill-health may threaten wealth through asset depletion to pay for medical and nursing care.

17.1.4 Organization of the Chapter

We begin by illustrating the strong positive relationship between health and income, using data from three countries—China, the Netherlands (NL), and the United States (USA)—that differ greatly with respect to level of development, economic inequality, labor market structures, and social welfare institutions. For each country, we show the contribution (in a purely statistical sense) that health differences make to income inequality, and, from the other side, the extent to which income variation accounts for health inequality. Having established the strength of the association between income and health, in sections 17.3 and 17.4, we turn to the mechanisms potentially responsible for the relationship and the extent to which it arises from a causal effect of health on income and vice versa. Section 17.3 identifies a number of routes through which health may impact income and wealth, paying particular attention to how economic inequality may be generated by health differences. The pathways considered are wages, work, human capital, early-life health risks, occupation, marriage, and medical expenditures. Evidence relevant to each broad pathway and more specific mechanisms is reviewed. Section 17.4 looks at the relation from the other direction: income (wealth) to health. Much of this discussion concerns whether income (wealth) has a causal impact on health over and above that of other socioeconomic characteristics, such as education and occupation, and after controlling for correlated determinants, such as time preferences and risk attitudes. Section 17.5 considers the logic and empirical support for the hypothesis that health is determined by economic inequality and by relative, as opposed to absolute, income. Finally, section 17.6 briefly summarizes the lessons that can be drawn from the literature about the nature of the income–health relationship and discusses what these imply for the normative evaluation of the distributions of income, health, and well-being.

17.2. HEALTH AND INCOME: A FIRST PASS

To whet the appetite, we illustrate the strength of the relationship between health and income in the USA, the NL, and China.[1] Our purpose is simply to show that there is a substantial and ubiquitous relationship that deserves attention and to assess its potential relevance to the explanation of inequalities in the distributions of income and health. The three countries are chosen primarily because of their differences. One is large, rich, and unequal, and it does not (yet) have universal health insurance coverage. Another is small, rich, and egalitarian, and it provides universal health coverage and extensive social protection, including DI, typical of northern continental Europe. The third is very large, much poorer (but rapidly becoming less poor), and less healthy than the other two, with increasing economic inequality and limited health and DI coverage. Differences in the wealth and economic structures of these countries, as well as their health and welfare institutions, might be expected to be reflected in the distributions of income and health, as well as the association between them.

Figure 17.1 illustrates the income gradient in self-assessed health (SAH) (Smith, 2004)—the most common survey measure of general health that invites a respondent to select one of four (China) or five (NL and USA) labels as the best description of his or her health. We focus on the percentage of individuals reporting less than *good* health, which always corresponds to the bottom two categories of SAH, by age-specific quartile groups of household per capita income.[2] In the USA, this percentage rises monotonically as income falls at all ages except among the oldest (70+). Even the poorest elderly, whose income should not depend on their current health, are more than twice as likely as their richest contemporaries to report less than *good* health. The pattern is similar in the Netherlands but for the absence of a gradient among young adults and a weaker gradient among the elderly. In both countries, the gradient increases until middle age and

[1] The US data are from the 2008 well-being module of the American Life Panel (ALP), which is nationally representative and implemented by RAND over the internet (https://mmicdata.rand.org/alp/). The Dutch data are from the 2011/2012 wave of the Longitudinal Internet Studies for the Social sciences (LISS), which is also nationally representative and has a similar protocol to the ALP (http://www.lissdata.nl/lissdata/). The Chinese data are from the 2006 wave of the Chinese Health and Nutrition Survey (CHNS) (http://www.cpc.unc.edu/projects/china/), which is representative of nine provinces (Heilongjiang, Liaoning, Shandong, Henan, Jiangsu, Hubei, Hunan, Guizhou, and Guangxi) that account for 41.7% of the total population of the country (National Bureau of Statistics of China, 2007). The sampled provinces are mainly located in the central-eastern, more developed and populated part of the country, although the eastern seaboard and the megacities located there are excluded. Clearly the CHNS is not nationally representative. By excluding both the poor western part of the country and the eastern seaboard, it likely understates the degree of economic and health inequality. Nevertheless, there is substantial variation in terms of GDP per capita and life expectancy across the provinces that are covered, and there is at least one province in each of the four economic regions of the country.

[2] We assign household per capita income to every household member and calculate the age-specific quartile groups of individuals.

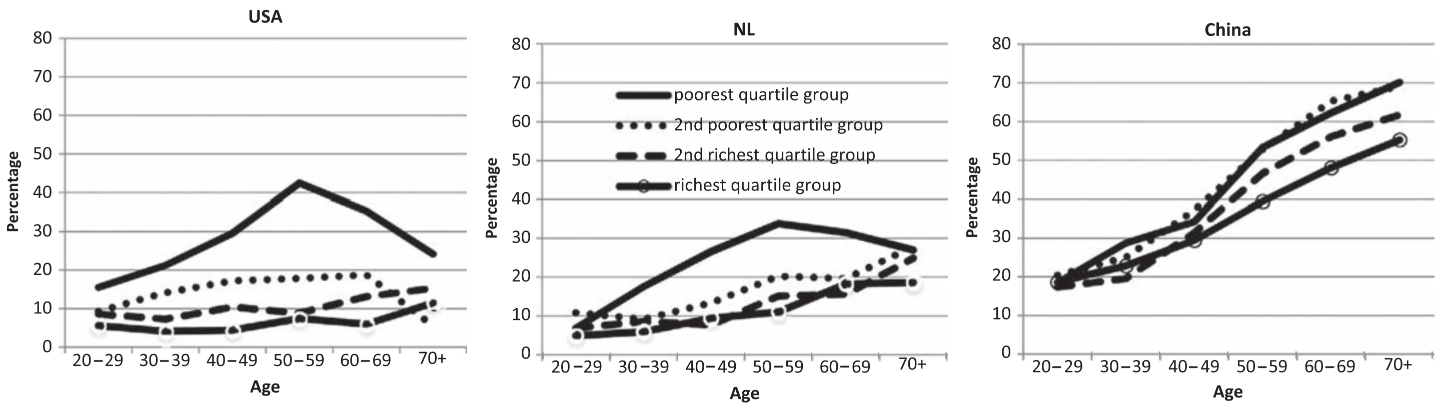


Figure 17.1 Percentage of individuals reporting less than *good* health by age-specific quartile groups of household per capita income in United States (USA), the Netherlands (NL), and China. *Notes: Authors' calculations from 2006 CHNS (China), 2011 LISS (NL), and 2008 ALP (USA). ALP and LISS respondents report health as being excellent, very good, good, fair (ALP)/moderate (LISS), or poor. CHNS respondents report health relative to others of their own age as very good, good, fair, or poor. Household income is before payment of taxes and Social Security contributions and after receipt of transfers. Income is annual for ALP and monthly for LISS and CHNS, with all incomes in local currencies. Per capita income assigned to each household member.*

narrows beyond that when retirement becomes more prevalent. This is consistent with employment being an important characteristic linking health to income. At the peak of the gradient between the ages of 50 and 59, more than 40% of the poorest Americans report their health to be less than *good*, compared with less than 10% of their richest compatriots. The inequality is narrower in the Netherlands, but the poor middle aged are still around three-and-a-half times more likely to report less than *good* health than are those in the top quartile group of the income distribution.

In China, the main health disparity is not between the poorest quartile group and the rest, as it is in the USA and the Netherlands, but, if anything, it is between the richest quartile group and those less privileged. There is no narrowing of the gradient in old age in China. In fact, health differences are greatest in the oldest age group, which is consistent with sizable inequalities in pension entitlements and health insurance coverage among the Chinese elderly but may also reflect the fact that the Chinese survey asks respondents to report health relative to others of the same age (see note to Figure 17.1).

Figure 17.2 shows the flip side of the relationship between individual health and income, with individual income measured by household per capita income. In all three countries, those in (at least) *very good* health have substantially higher incomes than those in *poor* health. In the USA at all ages, mean household per capita income falls as health drops from one category to the next. The health gradient in income peaks in the prime years of working life (40–49), when the mean income of those reporting *excellent* or *very good* health is around three-and-a-half times greater than the income of those in poor health. Even in old age, those with the best health have almost twice the incomes of those with the worst health. The health-related income gaps are narrower in the Netherlands. Even at the ages (50–59 years) where the disparity is greatest, those in the best health do not receive twice the income of those in the worst. The relative income differences by health in China are similar in magnitude to those in the Netherlands, except in old age when the gap widens, rather than narrows.

The individual health–income relationship remains strong after controlling for some potential correlates of both. Table 17.1 shows estimates from least squares regressions of the logarithm of household equivalent income of each individual for SAH, gender, age, ethnicity, education, and region (USA and China).[3] Conditional on these characteristics, in the US sample, the mean income of those reporting *very good* or *excellent* health is approximately 66–69% greater than that of someone reporting *poor* health (left panel, first column). This is larger than the relative income difference between those with the middle (post-high school vocational) and lowest (high school graduate or less) level of education, although it is not quite as large as the difference between university graduates and those with no more than high school education. Controlling for employment status has a very large impact on income differences by health. In the USA, the mean income of those

[3] Definitions of household equivalent income and the covariates are provided in Table A1.

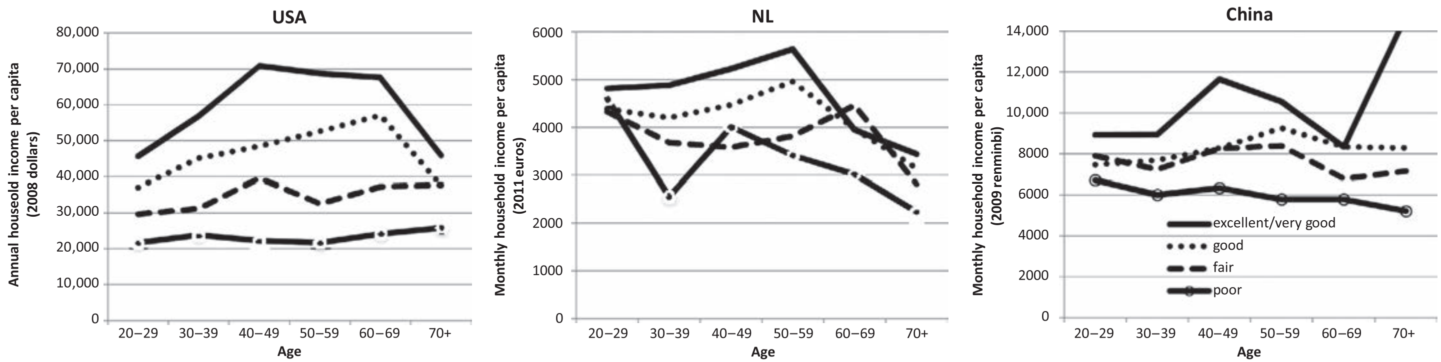


Figure 17.2 Mean household per capita income by self-assessed health and age in the United States (USA), the Netherlands (NL), and China.
Notes: As in Figure 17.1, the unit of analysis is the individual.

Table 17.1 Least squares regressions of log household equivalent income and decompositions of income inequality (relative Gini) in the United States, the Netherlands, and China

	Without control for employment status						With control for employment status					
	United States		Netherlands		China		United States		Netherlands		China	
	OLS coeff.	Inequality contribution	OLS coeff.	Inequality contribution	OLS coeff.	Inequality contribution	OLS coeff.	Inequality contribution	OLS coeff.	Inequality contribution	OLS coeff.	Inequality contribution
Health (SAH) (reference = poor)												
Moderate/fair	0.208**	6.5%	0.146**	3.6%	0.197***	3.0%	0.013	4.0%	0.092	2.8%	0.184***	2.5%
Good	0.508***		0.235***		0.244***		0.206***		0.152**		0.221***	
Very good	0.688***		0.314***		0.389***		0.334***		0.232***		0.362***	
Excellent	0.663***		0.369***				0.316***		0.268***			
Education (reference = low)												
Middle	0.445***	18.0%	0.069***	14.2%	0.370***	13.9%	0.390***	15.2%	0.040**	12.1%	0.314***	11.1%
High	0.882***		0.344***		0.754***		0.764***		0.305***		0.641***	
Gender (reference = female)	0.138***	1.8%	0.060***	1.5%	0.021	0.3%	0.103***	1.2%	0.036**	0.8%	−0.042*	0.4%
Ethnicity (reference = minority)	0.449***	6.7%	0.062**	0.4%	0.078**	0.6%	0.392***	5.6%	0.034	0.2%	0.051	0.3%
Age (reference = 20–29 years)												
30–39 years	0.195***	7.8%	−0.009	4.8%	0.048	2.5%	0.158***	8.0%	−0.060*	4.3%	−0.020	3.0%
40–49 years	0.392***		0.023		0.097**		0.354***		−0.036		0.023	
50–59 years	0.496***		0.161***		0.106**		0.513***		0.127***		−0.017	
60–69 years	0.585***		0.081**		−0.059		0.617***		0.132***		−0.247***	
70+ years	0.474***		−0.032		−0.098*		0.508***		0.041		−0.311***	
Region		3.6%		NA		8.0%		3.0%		NA		6.0%
Employment status (reference = employed)												
Unemployed							−0.782***	11.1%	−0.322***	7.7%	−0.527***	8.2%
Disabled							−0.892***		−0.248***		−0.361**	
Retired							−0.259***		−0.175***		0.401***	
Not working							−0.440***		−0.244***		−0.345***	
Unexplained (OLS residual)		55.7%		75.6%		71.8%		51.9%		72.0%		68.1%
Relative Gini		0.456		0.292		0.472		0.456		0.292		0.472
Number of observations	5050		4137		7694		5050		4137		7694	

Notes: Unit of analysis is the individual. Household equivalent income allocated to each individual in the household aged >19. Column headed "OLS coeff." gives coefficients from least squares regression of log household equivalent income on individual characteristics. "Inequality contribution" is the estimated contribution of the factor to inequality across individuals in household equivalent income computed from the Shapley value decomposition of the relative Gini index. SAH = self-assessed health. Definitions and means of the dependent variables and covariates are provided in Table A1. Reference category for ethnicity is not belonging to the main ethnic group (white (USA), Dutch (NL), and Han (China). Coefficients of regions (USA) and provinces (China) are not shown to save space. Region identifiers are not made available with the Dutch data. *, **, *** indicate significance at 10%, 5%, and 1% levels.

with at least *very good* health relative to those in *poor* health is reduced by half to 32–33% (right panel, first column). A large part of the strong relationship between income and health appears to be mediated through employment.

Multivariate analysis confirms what is suggested by Figure 17.2—that income gaps by levels of health are narrower in the Netherlands and China than in the USA. Without being conditional on employment, the Dutch reporting at least *very good* health have incomes approximately 31–37% higher than their compatriots in *poor* health. This is much lower than the respective relative disparity in the US sample. But being conditional on employment has a much smaller impact than it does in the USA, reducing the difference by around a quarter to 23–27%, which is only slightly less than the relative income difference of 30% between those with the highest and lowest levels of education. The more modest effect of being conditional on employment may be a reflection of the more generous DI in the Netherlands, which is evident in the coefficients on DI status. Being conditional on employment has little or no impact in the Chinese sample. Those reporting *very good* health have incomes approximately 36% higher than those in *poor* health. This is because employment differs less by health in China (see Figure 17.4).

Large differences in income by health do not necessarily imply that health statistically explains, let alone causally determines, a substantial part of income inequality. Whether it does depends on the degree of health variation that exists in the population, in addition to its partial correlation with income. The percentage of respondents reporting *poor* health is only 1.5% in the Dutch sample, rising to 3% in the USA and to 7% in the Chinese

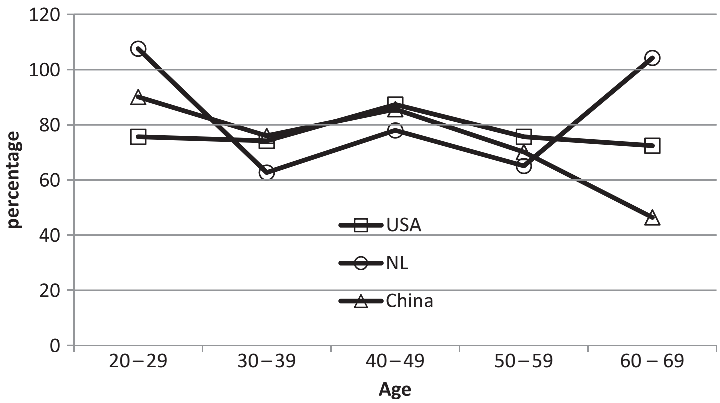


Figure 17.3 Mean earnings in bottom two categories of self-assessed health as a percentage of mean earnings in the top two categories, for China, the Netherlands, and the USA. *Notes: Authors' calculations from CHNS 2006 (China), LISS 2011 (Netherlands), ALP 2008 (USA). Samples restricted to those in work. Earnings include gross earnings/salary income from employment and profits from self-employment in the past year in the Netherlands and the USA, and gross wages, including bonuses and subsidies, in the last month in China. Self-assessed health is reported from five categories in NL and USA and from four in China. See notes to Figure 17.1.*

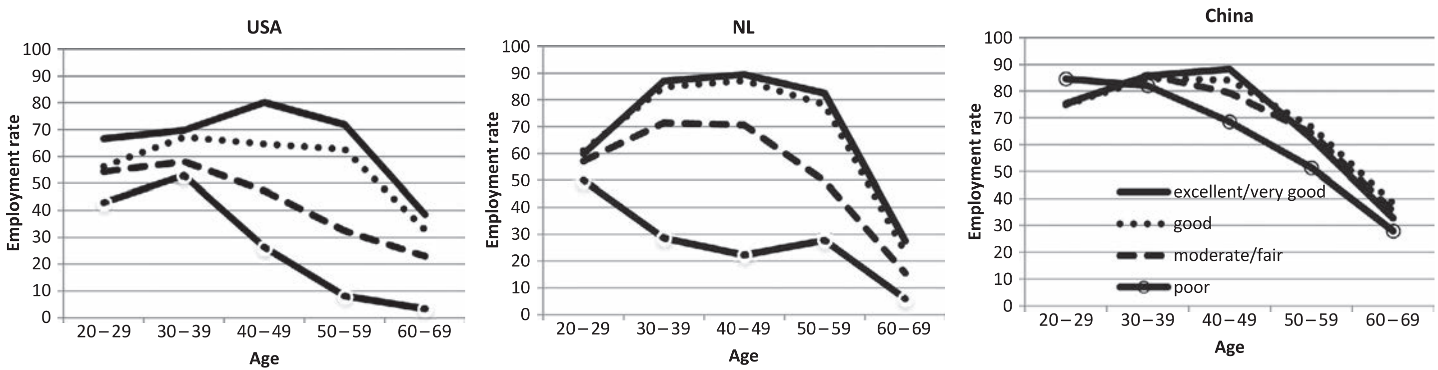


Figure 17.4 Employment rates by self-assessed health and age in China, the Netherlands, and the USA. *Notes: Authors' calculations from 2006 CHNS (China), 2011 LISS (NL), and 2008 ALP (USA). Employment includes full- and part-time work. For details of SAH by country, see notes to* Figure 17.1.

(Table A1). Differences in income between those with *poor* and higher levels of health may therefore make only modest contributions to the explanation of income inequality.

To give an impression of the contribution that health can make to the explanation of income inequality, we apply a simple version of a Shapley value approach (Sastre and Trannoy, 2002; Shorrocks, 2013) to decompose the relative Gini index estimated from the three datasets. This involves writing the income of each individual as the predicted value from the regression, plus the residual, and calculating the marginal impact on the Gini of neutralizing a variable by fixing its value across all individuals. This impact will vary depending on the covariates that have previously been held constant. The contribution of a variable to inequality is given by its average marginal impact across all possible sequences of neutralizing the set of all covariates.[4]

As would be expected, the estimated Gini indices reveal inequality to be lowest in the Netherlands (0.29) and of similar magnitude in the USA (0.46) and China (0.47). The percentage contributions of the factors to income inequality are reported adjacent to the respective column of regression coefficients in Table 17.1. Without conditioning on employment status, SAH explains 6.5% of income inequality in the US sample (left panel, second column). This is similar in magnitude to the contribution of race, a little less than that of age, and about one-third of that of education. The decomposition obviously depends on the specification of the regression model. Once employment status is added to the controls, health only explains 4% of income inequality, which is half of the contribution of age and a little more than one quarter of that of education.

Without conditioning on employment, health explains 3.6% and 3.0% of income inequality in the Netherlands and China, respectively. In each country, this is substantially more than the contribution of both gender and ethnicity. The health contribution is around 20% higher than that of age in China but less than the age contribution in the Netherlands. In both countries, as in the USA, the contribution of variation in SAH to the explanation of income inequality is substantially less than that of education. As predicted by the estimates of the regression models, controlling for employment status has less impact on the extent to which health explains income inequality in the Netherlands and China than it does in the USA.

Without conditioning on employment, which is the most obvious route through which health influences income, variation in SAH accounts for 6.5%, 3.6%, and 3.0%

[4] Income of individual i is given by $\exp\left(\hat{\gamma}_0 + \sum_{k=1}^{K} x_{ki}\hat{\gamma}_k + \hat{e}_i\right)$, where $\hat{\gamma}_k$ are coefficients from the OLS regression of log income and $\hat{e}_i$ is the residual. Inequality in this measure is computed and compared for different combinations of the regressors (x_{ki}) and residuals fixed at particular values. We compute the aggregate contribution of a factor, such as SAH, that consists of several categories represented by dummy variables, and we do not attempt to establish the contribution of each separate category. Use of the relative Gini and a regression model of log income ensures that the decomposition is insensitive to the values at which the regressors are fixed and to the estimate of the constant ($\hat{\gamma}_0$).

of income inequality in the USA, the Netherlands, and China respectively. Although these contributions appear modest, one must bear in mind that most inequality remains unexplained by an admittedly rather restricted set of factors in all three countries. SAH accounts for almost 15% of the explained income inequality in the USA and the Netherlands and just over 10% in China. In addition, SAH is only one measure of health and varies only over four or five categories. It is inevitable that such a variable will not be able to account for a substantial proportion of the variation in continuous incomes. Differential reporting of health may also result in SAH understating the correlation between income and health (Bago d'Uva et al., 2008). While it is established that SAH is an informative summary measure of health, the addition of more health measures, particularly continuous ones and those capturing disabilities, to the decomposition analysis would inevitably increase the proportion of income inequality that is explained by health variation.

Based on the regression and decomposition analysis presented, nothing can be inferred about causality—its direction or even existence. One can just as well look at health differences that are explained by income variation. Table 17.2 presents estimates from interval regressions of transformations of SAH on household equivalent income and the same covariates used in the income regressions (Doorslaer and Jones, 2003).[5] This health measure lies between 0 (minimum health) and 1 (maximum health). In the USA, the difference in predicted health between the richest and poorest quartile groups of household equivalent income is about 1.8 times greater than the difference between the youngest and oldest age group and between the highest and lowest education categories (left panel, first column). As could be anticipated from the narrower income gaps by health in the Netherlands, observed in Table 17.1, the health differences by income are also smaller than those in the USA (Table 17.2, middle panel). Still, the health disparity between the richest and poorest income groups is more than twice the difference between the top and bottom education groups. In China, the health differences by income are also substantially larger than the differences by education.[6]

Table 17.2 also presents results from Shapley value decompositions of inequality in transformed SAH. There are two notable differences from the decomposition of income

[5] Thresholds separating different categories of SAH are taken from external data. For China, we use the Chinese visual analog scale estimated from the World Health Organization Multi-Country Survey on Health and Responsiveness (Üstün et al., 2003), which gives thresholds 0 (minimum health), 0.50, 0.80, 0.91, and 1 (maximum health) corresponding to four categories of SAH. For the USA and the Netherlands, we obtain the thresholds 0, 0.428, 0.756, 0.897, 0.947, and 1 from the Canadian Health Utility Index (Feeny et al., 2002), with those thresholds dividing five categories of SAH.

[6] Conditioning on employment status reduces the health difference between the richest and poorest by about two-fifths in the USA. Doing so results in a smaller reduction in the income gradient in health in the Netherlands, which remains about twice the education gradient. Conditioning on employment has little impact in the Chinese sample.

Table 17.2 Interval regressions of self-assessed health (SAH) and decompositions of inequality (absolute Gini) in predicted SAH for the United States, the Netherlands, and China

	United States				Netherlands				China			
	Regression coeff.	Inequality contribution			Regression coeff.	Inequality contribution			Regression coeff.	Inequality contribution		
	Baseline	Baseline	With employment	With employment and health behavior	Baseline	Baseline	With employment	With employment and health behavior	Baseline	Baseline	With employment	With employment and health behavior
Income quartile group (reference = poorest 25%)												
2nd poorest	0.038***	44.0%	21.7%	14.3%	0.016***	34.5%	17.7%	13.8%	0.010**	10.5%	9.2%	8.8%
2nd richest	0.057***				0.024***				0.022***			
Richest	0.075***				0.036***				0.030***			
Education (reference = low)												
Middle	0.020***	25.9%	18.7%	9.6%	0.009**	15.9%	9.4%	6.4%	0.006	3.1%	2.6%	2.5%
High	0.042***				0.015***				0.011**			
Gender (reference = female)	0.001	0.2%	0.5%	0.1%	0.005*	3.5%	2.5%	1.9%	0.024***	8.6%	7.2%	7.4%
Ethnicity (reference = minority)	0.012***	4.4%	2.9%	1.9%	0.011**	3.1%	1.7%	1.3%	−0.002	0.1%	0.1%	0.1%
Age (reference = 20–29 years)												
30–39 years	−0.012**	20.4%	7.8%	4.4%	−0.021***	43.1%	26.3%	19.4%	−0.016***	63.8%	57.9%	55.9%
40–49 years	−0.035***				−0.035***				−0.041***			
50–59 years	−0.051***				−0.051***				−0.083***			
60–69 years	−0.054***				−0.048***				−0.111***			
70+ years	−0.042***				−0.055***				−0.148***			
Region		5.2%	4.8%	3.5%		NA	NA	NA		13.9%	13.1%	12.3%
Employment status			43.7%	34.2%			42.5%	37.9%			9.9%	10.0%
Health behavior			31.9%				19.2%				3.0%	
Absolute Gini		0.024	0.028	0.033		0.014	0.018	0.020		0.030	0.031	0.031
Number of observations	5050				4137				7694			

Notes: Unit of analysis is the individual. Column headed "Regression coeff." gives the coefficient from the interval regression of SAH with thresholds of categories as defined in footnote 5. "Inequality contribution" is the estimated percentage contribution of the factor to inequality in predicted health (obtained from the respective interval regression) computed from the Shapley value decomposition of the absolute Gini index. The latter takes values between 0 and 0.25, with those bounds indicating minimum and maximum health inequality, respectively. Definitions and means of the dependent variables and covariates are provided in Table A1. Reference category for ethnicity as in notes to Table 17.1. Region/province coefficients not shown. Regression coefficients are shown only for the baseline specification. Extended specifications sequentially add employment status (as Table 17.1) and proxies for health behaviors—smoking and weight. Smoking is measured by an indicator of ever having smoked in the USA and NL and currently being a smoker in China. Weight is summarized by a dummy for normal or overweight (18.5 < BMI < 27.5 for China and 18.5 < BMI < 30 for the USA and the Netherlands) and another for obesity (BMI > 30 for the USA and the Netherlands, and BMI > 27.5 for China). For underweight, BMI < 18.5 is the reference. *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively.

inequality. First, the categorical nature of SAH, modeled by interval regression, means that unexplained variation in health cannot be captured by the decomposition. All contributions refer to the percentage of the explained variation that is accounted for by a factor. Second, the absolute, rather than relative, Gini index (Yitzhaki, 1983) is a more appropriate measure of inequality in a bounded variable such as transformed SAH (Erreygers, 2009; Lambert and Zheng, 2011).[7]

Income quartile groups account for 45% of the explained inequality in SAH in the USA (left panel, second column). This is roughly equal to the contributions of age and education combined. Further evidence that employment is central to the association between health and income in the USA is provided by the fall in the income contribution by more than half, such that it becomes only slightly more important than education in explaining health inequality when employment status is added to the interval regression of SAH and so the decomposition (left panel, third column). Adding proxies for health behavior, in the form of indicators of weight and smoking, reduces the income contribution by about one-third more, which is suggestive of differences in lifestyle being an important reason why health differs by income.

In the Netherlands, income variation accounts for 35% of the explained inequality in SAH, which is more than twice the contribution of education. Adding employment status reduces the contribution of income by half, but it remains roughly twice that of education. Entering health behavior into the decomposition has a more modest impact. In China, income differences account for much less of the explained variation in health—around 9% irrespective of whether employment or health behavior is controlled for. As is also evident in Figure 17.1, health inequality in China appears to be driven mainly by age, which is perhaps surprising given that only the Chinese survey asks respondents to report their health relative to that of someone of the same age. On the other hand, a very steep decline in health with age in China would be anticipated from the lack of health insurance coverage, particularly for the elderly in rural locations, at the time of the survey, as well as from events experienced in the twentieth century.

Our empirical illustration demonstrates that there is a strong relationship between income and health. Income differences by health are large. Correspondingly, health disparities by income are wide. The relationship is stronger in the USA than in the Netherlands and China. The dissimilarities of the latter two countries imply that the strength of the relationship is not explained by a simple factor, such as the informality of the economy, universality of health insurance coverage, or generosity of welfare, but is likely a product of many such factors. Employment status, particularly in the USA among the three countries examined, is key to the relationship between income

[7] Given that the interval regression model has an additive specification and we use an inequality index that is invariant to equal additions, neither the constant nor the values to which factors are set when neutralized will affect the decomposition results.

and health. But it is unlikely to be the only mechanism. Even after controlling for employment, income differs greatly by health, and thus, health varies substantially with income.

A single health variable, self-assessed health, accounts for 6.5% of total income inequality and 14.6% of explained inequality in the USA. These estimates suggest the distribution of health in the population has a potentially important, although not central, role in explaining income inequality. But one could equally claim that variation in economic circumstances is key to the explanation of the health distribution. Indeed, income differences explain almost half of the inequality in predicted (self-assessed) health in the USA.[8]

While descriptive multivariate analysis and decompositions are useful in determining the strength of relationships, they tell us nothing about causality. Nevertheless, we hope to have convinced the reader that the association between income and health is sufficiently strong and pervasive such that it deserves to be probed by those seeking a better understanding of the distribution of income, as well as others aiming to account for disparities in health. In the next two sections, in turn, we consider the impact that health has on the distribution of income and the effect that income has on the distribution of health.

17.3. HEALTH DETERMINATION OF ECONOMIC INEQUALITY

17.3.1 Overview

How might the distribution of health determine the distribution of income? The most obvious effect is through physical and mental capacity for work. We begin this section by considering how health may impact productivity and wages. We then turn to the relationship between health and the quantity of work. Labor supply may be reduced at both the extensive margin, with illness in middle age tilting the balance in favor of early retirement, and the intensive margin, with part-time work becoming a more attractive proposition for some with a debilitating illness. In high-income countries, the employment effect on earnings will be directly cushioned by DI, but resultant moral hazard will indirectly contribute to the effect. Earnings losses may be exacerbated by discrimination, but legislation designed to prevent this may increase the impact on employment by constraining wage flexibility. Labor supply may be constrained by current sickness and influenced over the longer term by any downward revision of life expectancy following the onset of a major illness.

In addition to the immediate impact of ill-health on earnings, there may be important lifetime effects operating through education, occupation, and marriage. Illness in childhood can constrain opportunities for education and its efficiency in producing knowledge and skills. With few education qualifications, a frail young adult will be limited in his or her choice of occupation. The less healthy may also be constrained in their choice of partners. If there is sorting, such that the least healthy marry others of less than

[8] Of course, income would explain a much smaller proportion of the greater inequality in actual health.

average health or remain unmarried, then the contribution of health to inequality in household incomes will be even greater than its contribution to inequality in individual incomes. Ill-health may not only directly constrain the work effort of the disabled person, but it may also indirectly affect the labor supply of the spouse, who may face conflicting demands on his or her time to both replace lost earnings and provide informal care.

Health may impact the distribution of wealth, both directly and indirectly, through income and the accumulation of savings. Because the latter effect is cumulative, the contribution of health to inequality in wealth should be even greater than that to inequality in incomes. The healthy, expecting to live for longer, may save a larger fraction of their higher incomes, further increasing wealth disparities by health. When health insurance is incomplete and medical care must be paid for from an individual's own resources, illness can directly deplete wealth.

In the subsequent subsections, we consider the potential effects of health on income through wages, work, human capital, occupation, and household formation or spousal earnings. In each case, we elucidate the potential mechanisms and evaluate the evidence relevant to establishing the extent to which health differences contribute to economic inequality.

17.3.2 Health and Wages

Individuals in poor health have substantially lower earnings than those in good health. Figure 17.3 shows mean gross earnings of employees in the bottom two categories of SAH as a percentage of the mean in the two top categories for China, the Netherlands, and the USA, as estimated from the data sources in Section 17.2. Although there are discrepancies at the youngest and oldest age groups, mainly due to the small sample of individuals working and in less than *good* health in the Netherlands at these ages, among those working between the ages of 30 and 60, the earnings of individuals in the worst health are 15–40% below those in the best health in the three countries. In the middle age range, the health gradient in earnings is largest in the Netherlands, which may reflect both qualification for partial DI, which acts as an earnings subsidy, and the high rate of part-time employment. The relative earnings of the least healthy individuals decline most rapidly with age in China, where lower pension coverage leaves many with little option but to continue working despite deteriorating earnings capacity.

17.3.2.1 Productivity

Deterioration in health will often decrease labor market productivity and earnings capacity. But the multidimensionality of health and variation in the capacities and skills required for different occupations are reasons to expect a great deal of heterogeneity in the response of wages to health problems. A mobility-impeding disability obviously constrains the productivity of a manual worker much more than it does that of an office worker. Technology, particularly that which enables remote work and reduces the need for commuting, is making the productivity of workers who draw more on their brains than their brawn even less contingent on physical functioning. But the productivity of such workers is dependent on retaining cognitive functioning and is potentially vulnerable to mental health problems. Thus, the estimated effects of health on wages, derived from measures of different dimensions of health for individuals with different demographic and occupation characteristics, should display substantial variability.

The relationship between productivity and physical health is likely to be strongest in developing countries where there is a preponderance of low-skilled manual work.[9] Identifying the economic returns from health investments has been the motivation for macro studies, revealing a strong positive correlation between economic growth and initial population health (Barro, 2003, 2013),[10] and micro studies of the relationship between wages and health across individuals in developing countries (Strauss and Thomas, 1998). Our focus is not on the impact of health on average income. Rather, we are interested in the extent to which the distribution of health affects the dispersion in incomes.

In low-income settings, the relationship between health and productivity can sustain, exacerbate, and, in theory, even generate inequality. The crux of the argument is that health constrains productivity, and wages provide the means, through nutrition, to sustain health. A negative shock to either health or wages can generate a downward spiral into a nutrition-based poverty trap (Dasgupta, 1993, 1997). The theory of nutrition-based efficiency wages (Bliss and Stern, 1978; Dasgupta and Ray, 1986; Leibenstein, 1957; Mirlees, 1975; Stiglitz, 1976) generates the prediction of increased inequality as a result of the interdependence of nutritional status and productivity. As nutrition rises above the critical threshold of physiological sustenance, the marginal increases in productivity rise substantially, and diminishing returns eventually set in. This nonconvexity results in involuntary unemployment because the savings in labor costs from employing the poorly nourished unemployed at lower wages would be more than offset by the resulting loss in productivity. Individuals with initially fewer assets, who can invest less in health, are more likely to be involuntarily unemployed (Dasgupta and Ray, 1986). The poor are more likely to be malnourished and sick, and because they are sicker, they are more likely to be unemployed and so fall into destitution. Thus, inequality is exacerbated.

The relevance of this theory has been challenged by the observation that one of its central assumptions—that poor households are constrained to spend almost all their resources on food—is inconsistent with the observed behavior of the poor (Banerjee and Duflo, 2011). Starvation is not perennially present even in very low-income

[9] Strauss and Thomas (1998) note that the wage elasticity with respect to height (an indicator of health status) was almost eight times larger in Brazil (in the mid-1970s) than it was in the USA (in the early 1990s).

[10] Barro (2013) maps an extension of the neoclassical growth model that incorporates health, in addition to human capital, and sets a resulting research agenda, including the examination of how health and health policies may impact the evolution of income inequality.

countries (Strauss and Thomas, 1998). Ill-health may, however, be a source of a poverty trap and a contributor to inequality. For the income- and asset-poor with little formal or informal insurance options, the loss of productivity and earnings precipitated by illness results in reduced consumption opportunities and consequently nutritional deprivation, which further harms health and constrains productivity. The economic impact of illness is likely to be greatest on poor individuals because their livelihoods rest most on their health and because they have fewer assets that can be used to protect consumption and maintain nutritional status when illness strikes. Even if health shocks were evenly distributed across the population, their differential effects would increase economic inequality.

There is a vast body of evidence concerning the impact of health and nutrition on productivity and wages in low- and middle-income countries. The main motivation for this research is to evaluate the case for investment in health and nutrition programs as an instrument of development policy. Reviewing this literature would take us well beyond the scope of this chapter. In any case, a number of reviews already exist (Commission on Macroeconomics and Health, 2001; Deolalikar, 1988; Schultz, 2005, 2010; Strauss and Thomas, 1998; Thomas and Frenkenberg, 2002). Strauss and Thomas (1998) conclude that there is no robust, consistent evidence from nonexperimental studies that ill-health reduces productivity and wages, although it does reduce labor supply. In interpreting this conclusion, one needs to bear in mind that the wage response to ill-health can only be studied among employees. This misses the large informal sector of the economy in which the productivity of self-employed, mainly agricultural, workers could be expected to depend on health (Dasgupta, 1997). Evidence of an effect of nutrition, as opposed to health, is stronger. Strauss and Thomas (1998) are convinced that the positive impact of nutritional status (height and body size) on wages and micronutrients (particularly iron) clearly raises productivity. Calorie intake, when accurately measured, is found to have a positive effect on wages, at least among those initially malnourished.

17.3.2.2 Discrimination

Not all disabilities impede productivity; at least, not in all occupations. Nonetheless, equally productive disabled individuals may be paid less than their able-bodied counterparts because they are perceived to be less productive, or simply because of prejudice.

Discrimination against the disabled, as against other minority groups, comes in two varieties. What economists refer to as *taste discrimination* would be more commonly recognized as prejudice, and it arises from a preference of employers, or other employees, to keep disabled workers at a distance. Since Becker (1957), this has been modeled as a marginal cost, on top of the wage, that a prejudiced employer incurs in employing a member of the minority group. Such an employer will only hire a disabled person at a wage below his or her marginal product. Whether this discriminatory behavior is sustainable in a competitive market depends on the prevalence of prejudice relative to the supply of disabled labor (Becker, 1957). Under competitive conditions, nonprejudiced firms can undercut their prejudiced rivals, and discrimination will be competed away (Cain, 1986), unless prejudice arises from customers (Kahn, 1991).

This model was developed with the central purpose of explaining and understanding the consequences of discrimination against ethnic minorities. Although some disabilities, or rather handicaps, may still carry a social stigma, most are unlikely to make others, or at least a majority of others, uncomfortable. Stereotyping, or *statistical discrimination* (Aigner and Cain, 1977; Arrow, 1973; Phelps, 1972), seems a more probable source of bias against disabled individuals. In making appointments and wage offers, it is optimal for an employer to supplement information obtained from a noisy signal of productivity, such as a test score or qualifications, with knowledge of the average productivity of a group to which the applicant is observed to belong. Presuming disabled individuals are, on average, less productive, a disabled person would be offered a lower wage than a nondisabled applicant who performed no better with respect to the assessment criteria.

This theory does not help us explain earnings differentials between the disabled and nondisabled over and above those attributable to productivity differences. But it can explain part of the reason for the productivity deficit. If test scores or qualifications are a noisier signal of productivity for the disabled, perhaps because the tests are designed to discriminate between able-bodied applicants, then employers will put less weight on these criteria and more on the observed disability. Faced with a lower return, this group will invest less in human capital. Inequality will be greater than it would be if employers where blind to disability status, or legislation successfully forced them to act as if they were.

Besides its inability to explain earnings differentials beyond those attributable to productivity, the relevance of statistical discrimination as an explanation for health-related wage differences depends upon the extent to which these differences exist across easily recognizable disabilities with known average productivity differentials. A blind man is easily recognized and categorized. Someone with a heart condition is not. Even if all health conditions were observable, perhaps because applicants were required to declare them, how much would employers know about even average productivity specific to them? Rather than responding to an immediately recognizable disability group, employers might have only a partial, perhaps unconsciously biased, understanding of the productivity implication of an incompletely comprehensible health condition.

Empirical identification of discrimination against the disabled is difficult because disability, being an impairment of functioning, will certainly reduce productivity in many jobs. Getting hold of data that make it possible to control for real differences in productivity, and so isolate wage differences attributable to discrimination, is a tall order. Studies that control for little or no differences in health (Kidd et al., 2000) cannot credibly claim to identify discrimination against the disabled. But controlling for impaired functioning while comparing wage differences between the disabled and nondisabled seems to be like asking to have one's cake and eat it. One approach is to concentrate on the wage differences between individuals with easily observed disabilities, such as blindness, paralysis, or loss of a limb, which may be more likely to evoke prejudice, and others who may be completely able-bodied or disabled by an unobservable condition, such as a back pain or heart problems. US data from the 1970s and 1980s reveal that one-third to one-half of the wage differential between these groups is unexplained by wage determinants, including a battery of health indicators intended to capture differences in functional impairments (Baldwin and Johnson, 1994; Johnson and Lambrinos, 1985). This finding is indicative of substantial discrimination only if the controls are sufficient to mop up any productivity differentials. DeLeire (2001) suggests another approach which involves assuming there are no unobserved productivity differences between individuals who report a health problem but no work limitation arising from this and others reporting no health problem. In that case, all of the wage difference between these two groups can be attributed to discrimination. This can be taken as indicative of the discrimination against those with a work-limiting disability under the further assumption that the degree of discrimination is independent of the productivity loss arising from disability. Under these assumptions, only 7–11% of the wage difference not explained by observable characteristics between US males with a work-limiting disability and those with no disability could be attributed to discrimination.[11] It is difficult to hazard a guess as to whether this estimate lies closer to the truth than the larger earlier one, given that both rest on rather strong assumptions.

17.3.2.3 Nonwage Costs and Nonpecuniary Benefits

The previous two subsections considered wage variation arising from health-related differences in (perceived) productivity. The employer was assumed powerless to correct productivity differentials. A richer model of the demand for disabled labor relaxes this assumption. The productivity of someone bound to a wheelchair is contingent on adjustments made to the workplace—ramps, elevators, adjustable desk, etc. Installation of such facilities involves incurring a fixed cost that pays off through raising the marginal product of disabled workers (Acemoglu and Angrist, 2001). Treating labor as a quasi-fixed factor, with the simplifying assumption that there are fixed costs associated with employing disabled but not able-bodied workers, and supposing that workplace modifications close the productivity gap between disabled and able-bodied workers, the wage paid to the former will be lower by the amount of the (discounted) fixed costs (Acemoglu and Angrist, 2001). Wage differentials need not reflect only productivity differences or discrimination. Even with perfect measures of productivity, an empirical test of discriminatory behavior would be difficult. A second implication of this model is that employers are likely to be particularly apprehensive about appointing disabled workers. Fixed costs incurred

[11] The analysis is done for 1984 and 1993. Jones et al. (2006) apply the same approach to UK data and also find a small discrimination effect.

up-front must be compared with expectations of future marginal products and wages. A risk-averse employer will opt for labor with a higher proportion of variable costs.

Health-related wage variation could also arise from a willingness of employees to trade wage for nonwage benefits. The onset of a chronic condition would be expected to increase the value attached to employer-provided health insurance (Currie and Madrian, 1999). A worker suffering from a long-term illness would be more likely to accept a wage cut, or to forgo a pay rise, for fear of not being able to obtain insurance in a better paying job. Note that such health-related wage differentials do not imply differences in well-being. The individual is choosing to accept a lower price for his labor in return for obtaining a lower price for health insurance. Nevertheless, this would be an additional mechanism through which health differences may contribute to inequality in measured income, at last in countries with employment-based health insurance.

17.3.2.4 Evidence

Theory identifies mechanisms through which ill-health may reduce wages. But how large is the effect? Is health-induced variation large or small relative to overall wage inequality? The usual econometric demons—selection, omitted variables, reverse causality and measurement error—hinder attempts to answer these empirical questions. Evaluation of the evidence largely comes down to assessing the extent to which these problems have been overcome or avoided. Ill-health is likely to be a major reason for labor-force withdrawal (see next section). Estimation of the impact of health on wages from a cross section of workers or a balanced panel of individuals in continuous employment will overlook those whose wage opportunities were reduced most by ill-health and decided to stop working. Correction of this selection bias requires modeling employment, in addition to wages, with health allowed to impact both. With panel data, there may also be health-related attrition: those experiencing a marked deterioration in health might be more likely to drop out of the sample. Recognizing that individuals can influence their health through lifestyle, for example, leads to the realization that the same unobservable factors, such as time preferences, risk attitudes, and schooling quality, that influence job choices and thus wages may also condition investments in health. If panel data are available, then differencing can be used to purge the time-invariant unobservables correlated with health, or efficiency gains may be sought by using averages of assumed exogenous time-varying covariates to instrument health (Hausman and Taylor, 1981). Neither of these solutions is sufficient to remove bias if there is direct dependence of health on the wage. This is implied by Grossman's seminal model of health determination (Grossman, 1972a,b), according to which the wage influences both the costs of, and the returns to, investments in health (see Section 17.4.1). Correcting or avoiding the threat of simultaneity bias requires identification from exogenous variation in health that does not arise from wage differences. Prices of medical care and, in a developing country context, the local disease environment have been used as instruments for measuring health (Strauss and Thomas, 1998). However, it can be difficult to find variation in prices that is not endogenous to the choice of medical care provider, and geographic variation may be a rather weak instrument (Currie and Madrian, 1999). Disease exposure is often correlated with weather and agricultural conditions that would be expected to impact wages directly.

Currie and Madrian (1999) provide a comprehensive review of the US evidence on the health impact on wages (and on labor supply) up to the end of the twentieth century. They note three main deficiencies in this evidence base. First, estimates are sensitive to the measure of health, and variability in the measures adopted impedes comparability across studies. We would add that this sensitivity does not merely reflect inconsistency in the measurement of health but is due to intrinsic heterogeneity in the effect depending on the nature of the health condition. Second, few studies attempted to correct for the potential endogeneity of health, and those that did relied on rather dubious exclusion restrictions. Third, most of the evidence available referred to white (US) males. Picking up from where Currie and Madrian (1999) left off, in the remainder of this subsection, we focus on the evidence published since 1999, using data from high-income countries. The latter two criticisms have, to an extent, been addressed in the more recent literature. The increasing availability of panel data, particularly on older populations that experience the most variability in health, as well as population level administrative data, has reduced reliance on instruments to deal with endogeneity. Although many studies still tend to focus on males, there are many exceptions, and the evidence comes from a wider spread of countries. All studies cited are summarized in Table 17.3 for evidence relating to the USA and Table 17.4 for studies that use data from European and other high-income countries.

A fixed effects estimate obtained from retrospective life history data collected in the first wave (1992–1993) of the US Health and Retirement Study (HRS) suggests that a work limitation lasting at least 3 months reduces the wage rate by 4.2% for males and twice that for females aged 50–60 (Pelkowski and Berger, 2004). Given that 7–9% of individuals in this age range report such a health condition, these estimates suggest that ill-health makes a substantial, though not dramatic, contribution to wage inequality. Using 25 years of longitudinal data from the US Panel Study of Income Dynamics (PSID), Charles (2003) obtains a fixed effects estimate (corrected for selection into employment) only half as large for men for a similar measure of ill-health experienced by almost one-third of the sample at some time during the panel. This would suggest a much more modest contribution of ill-health to wage inequality. The lower estimate obtained by Charles may be attributable not only to the use of panel data, rather than retrospective, but also to estimation using a younger sample. An analysis of the same dataset and health measure, taking account of simultaneity as well as selection and unobservable heterogeneity, finds that below the age of 35 and above the age of 62 there is little difference between the wage profiles of individuals in good and bad health

Table 17.3 US evidence of health effects on labor market outcomes

	Study details						Effect of ill-health on			
Authors	Data	Sample	Health measure[a]	% in ill-health	Biases addressed	Estimator	Wage[b]	Employment[c]	Hours[d]	Earnings/income
Bound et al. (2010)	HRS 1992–1998 (4 waves)	Single men 50–62 years (baseline)	Latent: SAH instrumented by ADLs		SEL, UH, ME (JB)	DP model by SML		1 SD ↓ from average health @ 60 years → Pr (labor force exit) ↑ 8 ppt		
French (2005)	PSID 1968–1997	20–70 years	Work-limiting physical impairment or nervous condition	30 years: 6% 70 years: 40%	SEL, UH, SIM	FE with SEL and MSM	<35 and >62 years: no effect 35–62 years: ↓ 8–17%	<38 years: no effect 62 years: ↓ 45 ppt 66 years: ↓ 20 ppt	<40 years: no effect >40 years: 20–27% ↓	
Smith (2004)	HRS 1992–2000 (5 waves)	50–62 years	Major/minor new diagnosis	Major (minor): 20% (30%) incidence over 8 years	UH	OLS FD		Major: ↓ 15 ppt Minor: ↓ 4 ppt		Major: ↓ $4000 Minor: ↓ $500 (household annual income)
Pelkowski and Berger (2004)	HRS 1992–3 (life history data)	50–62 years	Work-limiting condition ≥ 3 months	Males: 8.7% Females: 7.3% (lifetime incidence)	SEL	Heckman selection	Males: ↓ 6.4% Females: ↓ 7.2%	Males: ↓ Females: ↓	Males: ↓ 6.3% Females: ↓ 3.9%	Males and females: ↓ 52% (lifetime earnings)
Charles (2003)	PSID 1968–1993	Men 22–64 years	Work-limiting physical impairment	31.6% (disabled at any time over panel)	SEL, UH	FE with Heckman selection	↓ 2%		↓ 6.7%[e]	↓ 15% (annual earnings)
Blau and Gilleskie (2001)	HRS 1992–94 (2 waves)	Men 50–62 years	SAH, work-limiting disability, major/minor diagnoses, ADLs		SEL, ATT, UH, SIM	FIML		Excellent → poor SAH: ↑ labor-force exit 5.7 ppt Work-limiting disability: ↑ labor-force exit 5.5 ppt		
Bound et al. (1999)	HRS 1992–1996 (3 waves)	50–62 years	SAH instrumented by ADLs (effect for SAH ≥ *good* → < *good*)		SEL, UH, ME (JB)	SML		Males: ↓ 55 ppt Females: ↓ 46 ppt		
McClellan (1998)	HRS 1992–94 (2 waves)	50–62 years	Major/minor new diagnosis, accident	Major: 3.5–6.3% Minor: 18.5–22% Accident: 5.3–8.7%	UH	OLS FD		Major: ↓ 17.5–26.3 ppt Minor: ↓ 1.8–5.1 ppt Accident: 0–↓ 2.1 ppt	[e]Major: ↓ 13.1–35% Minor: 0–↓ 3.8% Accident: ↓ 4.3–9.6%	

Notes: Table excludes earlier (pre-1998) studies summarized in tables presented in Currie and Madrian (1999). JB, justification bias; ME, measurement error; SIM, simultaneity bias; SEL, selection bias; UH, unobservable heterogeneity (OVB bias). See Table A2 for explanation other acronyms, including those of datasets and variables.

[a]Effects on outcomes are with respect to this health measure.

[b]Impact on hourly wage for those in employment.

[c]Impact on probability of employment unless otherwise stated.

[d]Relative impact on annual work hours.

[e]Assuming average annual work hours of 1800.

Table 17.4 European and rest-of-world evidence of health effects on labor market outcomes

	Study details							Effect of ill-health on		
Authors	**Country**	**Data**	**Sample**	**Health measure**[a]	**% in ill-health**	**Biases addressed**	**Estimator**	**Wage**[b]	**Employment**[c]	**Earnings/income**
García Gómez et al. (2013)	Netherlands	Administrative 1998–2005	18–64 years	Urgent, unscheduled hospital admission ≥3 nights	0.85%	UH, ME (JB)	DID and matching		↓ 7.1 ppt	↓ 4.8% (in work ↓ 2.9%, on DI ↓ 32.7%) (income)
Halla and Zweimuller, 2013	Austria	Administrative 2000–2007	Private employees 25–50 years	Commuting accident ≥1 sick day	0.67%	UH, ME (JB)	DID and matching	↓ 1.4% (daily wage)	↓ 3.3 ppt	
García-Gómez (2011)	9 EU countries	ECHP 1994–2001 (8 waves)	16–64 years	ΔSAH = top 2 → bottom 3, onset chronic illness/ disability		UH	DID and matching		ΔSAH: ↓ >5 ppt in 5/9 countries Chronic: ↓ >4 ppt in 6/9 countries	
Jackle and Himmler (2010)	Germany	GSOEP 1995–2006	18–65 years	ΔSAH = top 2 → bottom 5	Males: 12.5% Females: 13.3%	SEL, UH, ME, SIM	Semykina and Wooldridge (2010)	Males: ↓ 4.8%. Females: no effect	Males: ↓ 0.5 ppt Females: ↓ 1.5 ppt[d]	
Brown et al. (2010)	UK	BHPS 1991–2004 (14 waves)	18–65 years	SAH instrumented by health problems and reported limited activity		SEL, ME (JB), UH (Mundlak)	2 stage: GOP of SAH → ML of wage and employment probability	No signif. effect	ΔSAH (very good/good → poor/very poor): ↓ 11 ppt	
Jones et al. (2010)	UK	BHPS 1991–2002 (12 waves)	50–60/65 years	Reported limited activity and SAH instrumented by health problems	Limited activity: Males: 15.6% Females: 13.9%	UH (RE) (ME (JB) when use instrumented SAH)	2 stage: GOP of SAH → ML of retirement hazard		Effects on retirement hazard: Limited activity – ↑3.5 (M), ↑5.8 (F)[e] ΔSAH (excellent → poor/ very poor): ↑ 4.9 (M), ↑ 7.2 (F)	
Lindeboom and Kerkhofs (2009)	Netherlands	CERRA 1993–95 (2 waves)	Male 43–63 years, employed, household head	Work limitation instrumented by health problems (HSCL)	14%	UH (RE), ME (JB), SIM			↓ (large relative to nonhealth effects)	
Cai (2009)	Australia	HILDA 2003	Males 25–64 years	SAH (top 3 → bottom 2)	10.3% (bottom 2 SAH)	SEL, SIM	FIML	↓17–20%		
Disney et al. (2006)	UK	BHPS 1991–98 (8 waves)	50–60/64 years	SAH instrumented by health problems and reported limited activity		UH (FE (logit) and RE (hazard)), ME (JB)	2 stage: ordered probit of SAH → FE logit		↓ (large relative to nonhealth effects)	

Continued

Table 17.4 European and rest-of-world evidence of health effects on labor market outcomes—cont'd

	Study details							Effect of ill-health on		
Authors	**Country**	**Data**	**Sample**	**Health measure**	**% in ill-health**	**Biases addressed**	**Estimator**	**Wage**	**Employment**	**Earnings/income**
							(or RE hazard) of employment			
García Gómez and Lopez-Nicolas (2006)	Spain	ECHP 1994–2001 (8 waves)	16–64 years	ΔSAH = top 2 → bottom 3	8.1%	UH	DID and matching		↓ 5 ppt	Earnings ↓ €1740 Personal income ↓ €1033 Household income ↓ €1927
Moller Dano (2005)	Denmark	Administrative 1981–2000	20–54 years	Road accident casualty admitted to hospital	1.4%	UH, ME (JB)	DID and matching		Males: ↓ 11.8 ppt Females: no effect	Earnings: M ↓ 12%, F no effect Income: no effect
Au et al. (2005)	Canada	CNPHS 1994–2001 (4 waves)	50–64 years	SAH instrumented with HUI3 or health conditions		ME (JB)	2 stage: Ordered Probit SAH → LPM of employment		1 SD ↓ health: Males ↓ 25 ppt Females ↓ 19–21 ppt	
Contoyannis and Rice (2001)	UK	BHPS 1991–96 (6 waves)	16+ finished schooling	SAH, GHQ (psychological ill-health)	SAH < *good*: Males: 16% Females: 19%	UH	Hausman and Taylor (1981)	ΔSAH (*excellent* → < *good*): M/F: no effect F in work ↓ 2.8% GHQ: M ↓, F no effect		
Kerkhofs et al. (1999)	Netherlands	CERRA 1993–1995 (2 waves)	Male 43–63 years, employed, household head	Work limitation instrumented by health problems (HSCL)		UH, ME (JB)	2 stage: FE of work limitation → ML of employment hazard		↓ (ill-health dominant effect on exit through DI)	
Riphan (1999)	Germany	GSOEP 1984–94 (10 waves)	40–59 years	ΔSAH = ↓ 5 points on 10-point scale	3.1%		Logit (employment), unconditional DID (earnings/income)		↓ 6 ppt	Earnings: growth ↓ 1.9 ppt Household income: growth ↓ 5.2 ppt

Notes: Table excludes studies published pre-1999. JB, justification bias; ME, measurement error; SIM, simultaneity bias; SEL, selection bias; UH, unobservable heterogeneity (OVB bias). See Table A2 for explanation other acronyms, including those of datasets and variables.

[a]Effects on outcomes are with respect to this health measure.

[b]Impact on hourly wage for those in employment.

[c]Impact on probability of employment unless otherwise stated.

[d]Estimates from one of estimators used to correct for selectivity bias. Estimates from other estimators vary greatly, reaching effects of 16–19% points.

[e]Despite the large impact of a health shock on the retirement hazard, few individuals experience such a shock, and early retirement is simulated to be only 11% above what it would be in the absence of the shocks incurred.

(French, 2005). But in the prime ages of working life, individuals with a work limitation can command wages around 8–17% below the wages of those in good health.

Using UK panel data and the Hausman and Taylor (1981) estimator, Contoyannis and Rice (2001) find a significant effect of psychological health, but not general SAH, on wages for males, but no interpretation of the magnitude of the effect is given. For females in full-time employment, there is a significant effect of SAH; moving from less than *good* health to *excellent* health is estimated to result in a rather modest wage increase of less than 3%.

A cross-section analysis of data on Australian men that attempts to allow for full simultaneity finds a large effect of SAH on wages (Cai, 2009). But the instruments used (health conditions and behavior) are of dubious validity, and exploitation of the panel dimension of the data using a fixed effects estimator results in a large reduction of the estimate and loss of its significance.[12] Jackle and Himmler (2010) also resort to instruments in order to deal with endogeneity arising from more than correlated time-invariant unobservables in their analysis of German panel data. They assume that past doctor visits determine health but not labor-force participation or wages conditional on this. The rationale is that past medical care is the investment response to previous health shocks and need not be correlated with current labor market outcomes given current health. This may be so, but it does not allow for the possibility that individuals visit a doctor to obtain a sick note to justify work absence. For males, a worsening in reported health from *excellent* to *poor* is estimated to result in a 4.8% drop in the hourly wage. For females, there is no significant effect.

Although differences in health indicators and estimators still make it difficult to compare estimates, we tentatively conclude that ill-health does reduce wages in high-income economies, but the effect is more likely modest than substantial. Because most studies estimate the wage response to ill-health while controlling for occupation, this conclusion refers to the degree to which the wage adjusts within a given job. A larger wage effect may arise through ill-health induced changes in occupation. We examine this effect in Section 17.3.5.

17.3.3 Health and Work

As would be expected, employment rates vary a great deal with health. This is illustrated in Figure 17.4 for China, the Netherlands, and the USA. In the two high-income countries, already in young adulthood, individuals reporting *poor* health are much less likely to be working than their contemporaries reporting better health. The difference in employment by health widens until middle age, after which early retirement begins to reduce

[12] The author attributes this to greater measurement error in health changes and failure of the fixed effects estimator to deal with correlated idiosyncratic errors. Alternatively, it could be that the IV estimate is upwardly biased by invalid instruments.

labor-force participation of even those in good health and the employment gap narrows.[13] The relationship between employment and health is different in China in two respects. First, there is little or no difference in employment by health in young adulthood. Second, while a gap opens up at older ages, it never becomes as wide as that observed in the Netherlands and the USA. This is partly an artefact of self-assessed health being reported in four, rather than five, categories in the Chinese survey. But this is not the whole story. In the USA around the age of 50, the employment rate difference between those reporting *fair* health and those reporting *excellent* or *very good* health is about twice as large as the difference between those reporting *poor* health and those reporting *very good* health in China at the same age. The relationship between employment and health is weaker in China. This is not what one would expect given the differences in the structures of the economies and the greater role of manual labor in China. It may be that the more generous social protection in the high-income countries allows individuals experiencing health problems to more easily withdraw from the labor market.

In this section, we consider a number of mechanisms through which health may impact on employment, including the incentive effects of DI.

17.3.3.1 Incapacity and Involuntary Unemployment

The impact of ill-health on work may seem obvious. If you are sick, you cannot work. For relatively short term, acute illnesses, this is a reasonably adequate description of the effect. But it is an effect that could only explain temporary interruptions to earnings and income. Any substantial contribution of the distribution of health to the distribution of income is unlikely to operate through short-term sickness. A few chronic medical conditions are completely incapacitating. But most reduce capacity for work to some degree. Ill-health may reduce productivity, but it is unlikely that the marginal product is pushed to zero in all possible jobs. The wage could fall below the level at which work is considered worthwhile, but that is a choice rather than a *fait accompli*.

This reasoning rests on the assumption that wages are perfectly flexible. Evidence of a moderate impact of health on wages (see Section 17.3.2.4) may reflect institutional constraints on wage flexibility. Equal pay and antidiscrimination laws typically make it illegal for employers to pay disabled workers less than their able-bodied colleagues doing the same job. This may succeed in constraining health-related wage inequality but exacerbate disparities in employment.

The 1990 Americans with Disabilities Act (ADA) compels employers to accommodate disabled workers through adjustments to the workplace and outlaws discrimination against the disabled in hiring, firing, and pay. In theory, the impact on employment of the

[13] It is surprising that the total employment rate at ages 50–59 in the Netherlands sample is higher than that in the US sample. This may partly reflect the difference in years, but it seems that the ALP is underestimating the rate in the USA.

disabled is ambiguous. The threat of legal action for discrimination in hiring would tend to increase employment, while increased accommodation and hiring costs would reduce employment. Acemoglu and Angrist (2001) argue that the negative effects are likely to dominate. Before and after legislation, trends in employment are consistent with this prediction (Acemoglu and Angrist, 2001; DeLeire, 2000). Employment of disabled individuals was also reduced immediately after the introduction of the UK Disability Discrimination Act in 1996, which imposed similar obligations on employers as the ADA, before it recovered somewhat (Bell and Heitmueller, 2009). Digging deeper into the effect in the USA, Hotchkiss (2004) reveals that it is not due to individuals always classified as disabled being more likely to leave employment or less likely to enter employment. Rather, it is due to nonparticipants in the labor market reclassifying themselves as disabled after the passing of the legislation. Whether antidiscrimination legislation makes it more difficult for disabled individuals to obtain work may still be an open question. What seems clear is that there is no evidence that major legislation makes it easier for disabled individuals to gain employment.

17.3.3.2 Disability Insurance

Any illness-induced reduction in the offer wage makes labor-force participation less financially attractive. If the decline in health is sufficient to qualify for DI, then financial disincentives to work are compounded. Qualification for DI is not unambiguous (Diamond and Sheshinski, 1995). It is typically not determined by the presence of a precisely defined medical condition but is assessed on the basis of the vague concept of "capability of performing paid work," perhaps taking account of workplace conditions and occupation. There is subjectivity in whether a person considers himself incapable of work, as well as whether the adjudication officer agrees. Financial incentives can tilt the balance in favor of applying for DI. For a given degree of work incapacity, withdrawal from employment is more likely when social protection is available to cushion the resulting loss of income (Autor and Duggan, 2006; Bound and Burkhauser, 1999; Gruber, 2000; Parsons, 1980).

By increasing the likelihood of labor-force withdrawal but compensating for the resulting income loss, the existence and generosity of DI may simultaneously strengthen the relationship between health and earnings, and it may weaken the relationship between health and income.[14] For a given distribution of health, income would be expected to be more equally distributed in countries with generous DI. But the equalizing effect of social protection may be weakened by a moral hazard effect that is stronger for individuals with lower earnings potential. If, as is usually the case, the replacement rate

[14] Hurd and Kapteyn (2003) find that income change is more sensitive to the level of self-assessed health of individuals aged 51–64 in the Netherlands, which has more generous DI, than in the USA. A possible explanation is that the moral hazard effect outweighs the income replacement effect of DI. Individuals in poor health are more likely to withdraw from employment in the Netherlands at prime working ages, consistent with what is observed in Figure 17.4.

is decreasing with predisability earnings, the financial disincentive to continue in employment will be greatest for lower-paid workers. As a result of these differential incentives, onset of a given disability is more likely to result in employment withdrawal and loss of earnings (only partially replaced by DI) of lower-paid workers. This differential moral hazard effect will tend to increase income inequality relative to the hypothetical situation in which the financial incentives arising from DI have no impact on employment. But in the complete absence of DI and the income protection it provides, income inequality would be likely to be even greater. In addition to the strength of the differential moral hazard effect, the extent to which DI reduces income inequality will depend on the incidence of disability. The equalizing effect will be greater if, as is likely, the poor are more likely to become disabled. The lower paid are more likely to both benefit from social DI and respond to its financial disincentives to work.

DI rolls have been rising over much of the past 30 years in many high-income countries (OECD, 2010; Wise, 2012). Steep downward trends in mortality rates, accompanied by a compression of disability, if not disease, into fewer years before death (Crimmins and Beltran-Sanchez, 2011; Cutler et al., 2013; Milligan and Wise, 2012), suggest that this is not because populations are becoming less healthy. Looser eligibility criteria and increased returns from claiming DI relative to those available from work are considered to be the chief culprits (Autor and Duggan, 2006; Bound and Burkhauser, 1999; OECD, 2010; Wise, 2012). We will not evaluate the evidence concerning the strength of the moral hazard effect of DI (Bound and Burkhauser, 1999) but concentrate on the proposition that it interacts with increased economic inequality arising from structural changes in the economy to reduce labor-force participation of low-skilled, low-paid workers (Autor and Duggan, 2003, 2006).

The falling relative wages and employment opportunities experienced by low-skilled workers in the USA and other high-income countries since the 1980s increases the attractiveness of DI for this group (Autor and Duggan, 2003; Black et al., 2002). Dependence of DI awards on ability to engage in gainful employment results in DI applications tending to rise in an economic downturn (Autor and Duggan, 2006) and suspicion that some governments deliberately use DI to disguise long-term unemployment. After the loosening of the Social Security Disability Insurance (SSDI) eligibility criteria in the USA in the mid-1980s, the sensitivity of applications to adverse economic conditions increased by at least twofold, and high school dropouts became twice as likely to exit the labor force on the occurrence of a negative shock to the economy (Autor and Duggan, 2003). The effect is compounded by the indexing of the SSDI benefit formula to average wage growth. As a result, the replacement rate has increased for individuals whose wage growth lagged the average, as has been the case for low-paid workers in the USA over the past 30 years (Autor and Duggan, 2003, 2006). The replacement rate was further increased by rising real expenditures on Medicare, the health insurance program for the elderly that SSDI beneficiaries are given entitlement to before reaching the age of 65. Taking into account these fringe benefits, the DI replacement rate for a 50- to 61-year-old male at the 10th percentile of the earnings distribution increased from 68% in 1984 to 86% in 2002 (Autor and Duggan, 2006). At the 90th percentile, the increase was much more modest, from 18% to 22%. As would be expected given these differential incentives, SSDI enrolment rates are much higher and have increased much more rapidly for low-skilled individuals. For male high school dropouts aged 55–64, the rate increased by 5% points between 1984 and 2004 to reach almost 20% (Autor and Duggan, 2006). The increase was only 1% point (to reach 3.7%) for the college educated.

In the USA, reduced earnings prospects relative to the average and increased DI replacement income relative to those reduced earnings, separately and in combination, have reduced the incentives for low-skilled workers with health problems to continue working. Increased reliance of this disadvantaged group on DI may be both a consequence of rising economic inequality and, given the replacement ratio is less than one, a contributor to it. This process may also operate in Europe, which also witnessed increased wage inequality and, at times, rising DI rolls since the 1980s.

17.3.3.3 Preferences

In addition to an indirect effect through DI, ill-health may shift the reservation wage directly by changing preferences for consumption relative to leisure. The direction of the effect is ambiguous. Ill-health would be expected to increase the disutility of work. But it may also reduce the marginal utility of a number of leisure activities, such as sport. The direction of the effect on the marginal utility of consumption is even more difficult to predict. A disabled person may derive less, or no, pleasure from some goods, including sports equipment or travel, but become more dependent on others, such as pharmaceuticals, heating, and private as opposed to public transport. Comparing the relationship between subjective well-being and consumption—proxied by permanent income—for older Americans with and without chronic illness, Finkelstein et al. (2013) infer that the marginal utility of consumption decreases with ill-health. If we assume that the increased disutility of work dominates, such that the marginal utility of leisure rises, this gives a clear prediction that the marginal rate of substitution of leisure for consumption rises when health falls. The reservation wage rises and labor-force withdrawal becomes more likely.[15] More direct evidence suggests that a work-limiting health problem is equivalent to aging around 4 years in the extent to which it increases the willingness of older (58+) Americans to trade consumption for leisure (Gustman and Steinmeier, 1986a,b).[16]

[15] The evidence from Finkelstein et al. (2013) is obtained from a sample of nonworking, elderly individuals, and thus, one cannot necessarily infer from this how ill-health affects preferences for consumption relative to leisure.

[16] For other estimates of health-specific utility function parameters in models of health and retirement, see Sickles and Yazbeck (1998) and Bound et al. (2010).

While these estimates support the plausible proposition that ill-health shifts preferences away from work, one should be careful not to overlook the multidimensionality of health. A physical disability may reduce the marginal utility of many leisure pastimes as much as, or more than, it raises the disutility of work. A chronic illness such as diabetes may have little or no impact on preferences for leisure relative to consumption. The empirical content of a prediction that preferences for work increase with a characteristic we conveniently refer to as *health* is blurred if that characteristic, or at least the means of measuring it, is not well defined. In the social sciences, we are rather attached to the concept of health, despite finding it difficult to define what we mean by it (Twaddle, 1974). Estimates of the impact of ill-health on work–leisure preferences are likely to vary with the dimensions of health examined and the indicators used to measure them.

17.3.3.4 Life Expectancy

Expected longevity is an additional mechanism through which differences in health may contribute to observed differences in income and wealth in a cross section. In the standard life cycle model of consumption with no bequest motive in which there is dissaving before death, a longer length of life is predicted to increase labor supply (and saving) at any given age (Chang, 1991; Hammermesh, 1984). This is basically a wealth effect. Increased lifespan implies greater lifetime potential income. The resulting increase in demand for consumption prompts a rise in labor supply.[17] We may refer to this as the *horizon effect*. Individuals in poor health work less because they do not have to provide for an extended old age. For a given degree of functional impairment, illnesses that are life-threatening, or at least life-shortening, should be observed to reduce earnings by more than chronic disabilities that present no threat to longevity (McClellan, 1998).

Recognizing that the length of life is uncertain and annuity markets are incomplete gives rise to an offsetting effect. Through the mortality risk on savings, or the prospect of dying before being able to enjoy the fruits of one's savings, variation in survival probability affects the marginal return on savings and, consequently, the marginal rate of substitution between consumption and leisure (Chang, 1991; Kalemli-Ozcan and Weil, 2010). Through this *uncertainty effect*, a reduction in the probability of death raises the return on and, thus, level of savings, making it possible to reduce labor supply, perhaps by retiring earlier, even when there is a longer expected length of life to be provided for. The chances of this *uncertainty effect* dominating the *horizon effect* decrease with the initial mortality rate (Kalemli-Ozcan and Weil, 2010). In high-income countries, one expects ill-health and reduced longevity to be associated with lower earnings (and wealth).

[17] Similarly, increased loss of time due to sickness within any period of life is predicted to reduce labor supply in that period, although increases in the expectation and uncertainty of future sickness are predicted to increase current work effort (O'Donnell, 1995).

17.3.3.5 Evidence

Apart from the ambiguous effects of preferences and longevity, theory gives the clear prediction that ill-health reduces work effort. Under reasonable assumptions for high-income economies, the effects of preferences and longevity go in the same direction. But what is the size of the overall effect of ill-health on employment?[18]

Most studies address this question using samples of older individuals. Effectively, they estimate the impact of health on (early) retirement. In general, health is found to be an important determinant of retirement (Currie and Madrian, 1999; Lindeboom, 2012). How important depends on the measure of health and estimator adopted, as well as the context.

Estimating the effect of health on employment is complicated by econometric obstacles similar to those confronted when trying to identify the impact of health on wages, only the measurement error problem becomes particularly thorny. Most studies have relied on survey data and self-reported measures of ill-health. One would expect the reporting of health to be endogenous to employment. Put most crudely, individuals who have decided not to work may lie about their capacity to work either to reduce the stigma attached to voluntary inactivity or because they are claiming DI. But the phenomenon need not be so blatant. The threshold of functioning at which an individual considers himself to be incapable of work may be influenced, possibly subconsciously, by financial incentives to work, job stimulus, length of working life, contact with others claiming DI, and so on. Reported ill-health may reflect motivations for not working. This so-called *justification bias* has been a major concern in the literature (Bound, 1991). The evidence tends to suggest that it leads to substantial overestimation of the effect of ill-health on employment (Bazzoli, 1985; Bound et al., 2010; Lindeboom, 2012; Lindeboom and Kerkhofs, 2009), although there are dissenting findings (Dwyer and Mitchell, 1999; Stern, 1989).

Replacing reported work incapacity with more objective indicators of chronic illnesses or future mortality reduces the risk of justification bias but increases classical measurement error and may result in underestimation of the effect (Bound, 1991). Instrumenting reported work limitations with more objective health indicators is arguably a better approach (Bound, 1991; Bound et al., 2010; Stern, 1989) but requires that the indicators are free of the justification bias, which may be a strong assumption given these are often also self-reported.

[18] We address this question in the context of high-income countries. In low-income countries with less formal labor markets, identification of the impact of health on employment and earnings is more challenging. The literature tends to focus on the extent to which households can smooth consumption over health shocks in the absence of formal health and disability insurance (Gertler and Gruber, 2001; Mohanan, 2013; Townsend, 1994). Reviewing that literature would take us beyond the scope of our objectives.

All studies from North America and Europe that attempt to deal with justification bias by treating self-reported health as endogenous confirm that health is an important determinant of labor-force participation (Au et al., 2005; Blau and Gilleskie, 2001; Bound et al., 1999; Brown et al., 2010; Disney et al., 2006; Dwyer and Mitchell, 1999; Jones et al., 2010; Kerkhofs et al., 1999; Lindeboom and Kerkhofs, 2009; Sickles and Taubman, 1986; Stern, 1989).[19] Health also emerges as a strong determinant of retirement in structural life cycle models of older US males (Bound et al., 2010; French, 2005; Sickles and Yazbeck, 1998). Bound et al. (2010) find that, although responding to justification bias by instrumenting self-reported health with objective indicators greatly reduces the health effect, this effect remains very large. Before the early retirement age, an older single male in bad health is five times more likely to withdraw from the labor force than an equivalent in good health.[20] When US workers reach 62, the age at which Social Security pensions can first be claimed and financial incentives shift in favor of retirement, the probability of withdrawal rises from 0.1 to 0.17 for those in poor health and from 0.025 to 0.049 for those in good health. In absolute terms, those in poor health react more to the financial incentives such that the difference in employment probabilities widens when it becomes financially more advantageous to retire. Ill-health and financial incentives interact. Financial incentives appear to tilt the balance toward considering a health problem to be incapacitating. From a low level of health, marginal deteriorations in health have a large effect. At age 62, when the health of a worker in poor health decreases by one-half of a standard deviation, the probability of that worker's labor-force exit increases from 0.17 to 0.27; this change is 10 times larger than the change in probability for a comparable worker in average health.

Extending the scope of analysis beyond older US males, French (2005) finds a very strong health effect that varies across the lifecycle. Below the age of 40, there is no difference in the employment rate of men in good and bad health. At 40, an incapacitating physical or nervous condition is estimated to reduce employment probability by 5% points for a stereotypical male. The effect reaches a peak of 60 points at age 58 before declining to 45 points at age 62 and 20 points at age 66. In responding to these very large effects, one needs to bear in mind that they are not corrected for justification bias. Despite the very large estimated health effect, the author argues that it is modest in terms of the proportion of the total decline in older male labor-force participation that can be attributed to population health. The percentage of males reporting an incapacity rises from 20% at age 55 to 37% at age 70. Applying the estimated health effect, this decline in health can explain only 7% of the 74% fall in labor-force participation between these ages. This

[19] All mentioned studies that are not included in the tables presented in Currie and Madrian (1999) are summarized in Table 17.3 for the USA and Table 17.4 for elsewhere.

[20] Health is modeled as a latent variable. "Good health" refers to an average score and "bad health" to a score one standard deviation below the average.

may be so, but the estimates of both French (2005) and Bound et al. (2010) imply that ill-health can account for a substantial fraction of labor-force withdrawal before the early retirement age of 62.

Establishing a substantial effect of health on employment only takes us part way toward our objective of gauging the contribution of health to the distribution of income. We need to know the earnings and, ultimately, income consequences of illness-induced loss of employment. Reviewing the earlier US evidence, Currie and Madrian (1999) conclude that ill-health has a large negative effect on earnings, which operates mainly through reduced hours of work, including nonparticipation, rather than reduced wages.

From the US HRS life history data, Pelkowski and Berger (2004) find that 7–9% of individuals over the age of 50 have experienced a work limitation at some time that, on average, is estimated to result in the loss of half of their potential future earnings. By far the greatest part of this substantial loss is from the reduced likelihood of working, which falls to around half of its counterfactual value. Taken at face value, these estimates suggest that health is a major determinant of earnings differences. Caution should be exercised in their interpretation, however. A major concern is the retrospective nature of the data, which may be vulnerable to recall, as well as justification, bias. There could be a tendency to report health events that did have labor market consequences and to recall changes in labor outcomes that coincided with periods of illness.

Charles (2003), using the US PSID, finds that annual earnings of initially employed individuals fall, on average, by 15% around the time of onset of disability. Given that almost one-third of the panel experiences a work limitation at least once, this suggests that ill-health is an important contributor to cross-sectional inequality in earnings. But earnings do not remain as depressed as they are immediately following the onset of ill-health. Some illnesses recede and disabilities can be adapted to through retraining. Within 2 years of the onset of disability, about half of the earnings loss is recovered. Subsequently, earnings continue to trend upward, rather slowly, toward the level at which they would have been without the worker having experienced the disability. These findings suggest that Pelkowski and Berger (2004) substantially overestimate lifetime earnings losses by extrapolating from the contemporaneous impact of ill-health on earnings.

While the PSID study provides valuable insight into the impact of ill-health on earnings, it is weakened by reliance on self-reported work limitation with no correction for justification bias. The steepest drops in earnings occur in the period between 1 and 2 years prior to the reporting of a work limitation. One interpretation is that health is declining and impeding labor outcomes prior to the point at which a health problem is reported. But another is that decreasing returns to and motivation for work lead to the reporting of a disability, perhaps to justify entry to DI.

Besides its vulnerability to justification bias, another limitation of the reported work capacity measure is that it focuses on functional impairment and does not discriminate other dimensions of health, such as longevity. This is not sufficiently recognized in part of the literature, which tends to presume that the ideal measure of health would be one that accurately informs of work capacity (Bazzoli, 1985; Bound, 1991; Lindeboom, 2012). Implicitly, the assumption is that there is only one mechanism through which health impacts employment—physical and mental capacity to perform work-related tasks. Recognizing the other mechanisms identified in the preceding subsections, health conditions that differ with respect to their implications for current functional impairment, prospects of recovery, survival chances, and so on can be expected to have differential impacts on work.

McClellan (1998) makes an interesting distinction between three types of health outcomes. *Major health events*, such as a severe heart attack or a stroke, imply both acute and long-term functional impairment and reduced life expectancy. Onset of a *chronic illness*, such as a heart condition or diabetes, does not dramatically affect current functioning but may affect labor supply through expectations given prospects of a degenerative disease and long-term impairment. At the other extreme, *accidents* have an immediate impact on functioning but are less likely to have any effect on preferences or health expectations in the long term. Using the first two waves of the US HRS, McClellan finds evidence consistent with the nature of the health event having an independent effect on employment over and above that of the degree of functional limitation arising from it. For a given change in functioning (measured by Activities of Daily Living (ADL)), the reduction in the employment probability is 40 points greater when it is precipitated by a major health event. The employment probability falls by 14% and 35% points for males incurring a new chronic illness with moderate and major reductions in functioning, respectively. Accidents are not significantly associated with a decline in employment. McClellan interprets this as indicative of employment effects being muted when the health event does not have consequences for long-term health expectations.[21]

Smith (2004) extends this type of analysis by using more waves of the HRS to look at longer term effects. Onset of a major health condition (i.e., cancer, heart disease, lung disease) is associated with an immediate reduction of 15% points in the employment probability of males among whom just over half were working at baseline. After 8 years, those contracting such an illness are 27 points less likely to be working. The short- and long-term effects for those succumbing to a minor chronic illness (i.e., hypertension, diabetes, arthritis, heart attack, angina, and stroke) are reductions of 4% and 11% points, respectively. The average loss in annual household income immediately following the onset of a major chronic illness is $4000, which rises to $6250 after 8 years. Cumulating these losses gives a total loss of income over 8 years of $37,000. The cumulative income loss associated with a minor condition is almost $9000.

[21] It could also be that the number of survey participants experiencing an accident does not provide sufficient power to precisely estimate the effect. Using population data, Moller Dano (2005) and Halla and Zweimuller (2013) do find significant effects of accidents on employment (see below).

Although this evidence of sustained reductions in employment and income appears to contradict that of Charles (2003), it does not once the older age of the HRS sample is recognized, and the heterogeneity in the effects estimated by Charles is scrutinized. Men that become chronically disabled, defined as reporting a work limitation in every period after the initial onset, experience an estimated initial earnings loss of 21% with little or no recovery over time. The initial loss is also greater and the recovery absent for older men. So, both studies find substantial permanent losses of income for older males succumbing to a chronic condition.

Bound et al. (1999) focus on the implications of the dynamic evolution of health for the continued labor-force participation of older (50+) US workers. They find that it is health deterioration, rather than the health level arrived at, that most affects labor-force exit. Employment is not simply dependent on current functional capacity because there can be adaptation to impairments. A drop to any given level of health implies lower expectations of future health that may have an independent effect on the decision to continue in employment. Disney et al. (2006) take the same approach with British data and find a different pattern of behavior. Controlling for past health, lower current health still increases the likelihood of job exit. But for any given level of current health, a lower level of past health also raises the probability of retirement. A possible interpretation of this result is that individuals take time to revise their health expectations downward. Only when health is persistently lower is the decision made to retire, which may be difficult to revoke.

Cross-country comparisons can be useful for exploring the extent to which the employment and income responses to ill-health appear to be influenced by employment and social policies. García-Gómez (2011) compares the likelihood of continuing in employment following a sustained drop in SAH in nine EU countries. She finds the employment effects are largest in Ireland, a country in which DI claimants are not allowed to engage in any form of paid work, followed by Denmark and the Netherlands, where replacement rates are highest and, as in Ireland, there is no quota on the percentage of employees that must be registered disabled. In France and Italy, the two countries that impose the highest quotas on disabled employees, there is no significant impact on employment. These findings suggest, but do not confirm, that the employment (and income) effect of ill-health is highly contingent on policies influencing both the demand and supply of disabled labor.

Rather than attempt to identify the health effect from data on self-reported general health, which is difficult to interpret and potentially endogenous, or reported work incapacity, which is possibly even more endogenous due to justification bias, three European studies have concentrated on more narrowly defined health events that are abrupt and unforeseen and so more plausibly exogenous to lifecycle planning of health and labor supply (García Gómez et al., 2013; Halla and Zweimuller, 2013; Moller Dano, 2005). Unlike structural models that aim to estimate the endogenously determined lifecycle profiles of health and labor supply (Bound et al., 2010; French, 2005; Sickles and Yazbeck, 1998), the objective of these studies is to exploit some unanticipated shift in the health profile to estimate the response of employment and income. Two of the studies use accidents (Halla and Zweimuller, 2013; Moller Dano, 2005), and the third relies on urgent and unscheduled hospital admissions (García Gómez et al., 2013). Implementation of this strategy is made feasible by the availability of population, or near population, data from administrative registers providing sufficient observations of relatively rare health events for which matches can be found from millions of control observations. Use of administrative records greatly reduces measurement error and avoids the justification bias that plagues estimates based on reported health. All three of the studies combine matching, which deals with observable differences, with taking difference-in-differences to eliminate correlation of the health event with time-invariant unobservables.

These studies consistently find that a health shock reduces the probability of employment. The estimates range from a 3.3% point reduction as a result of a commuting accident in Austria (Halla and Zweimuller, 2013), through a 7.1 point drop due to an acute hospital admission in the Netherlands (García Gómez et al., 2013), to a 11.8 point fall following a road accident experienced by men in Denmark (Moller Dano, 2005).[22] That the estimates differ in magnitude is to be expected given the narrow definition of the health events from which they are identified. Commuting accidents mostly give rise to musculoskeletal impairments, and diseases of the circulatory, digestive, and respiratory systems are all important causes of acute hospital admissions. Such different conditions would be expected to have different effects. A focus on specific health events that occur suddenly enhances the internal validity of these studies, but generalization to other forms of health deterioration cannot be presumed. There is no avoiding the multidimensionality of health and the consequent heterogeneity in its effects.

These studies confirm that ill-health causes employment to fall, a conclusion that could be made with less certitude from more weakly identified estimates. A less predictable finding is that the effect is persistent. All three studies find that the probability of employment remains reduced by a health shock for at least 5 years following its occurrence. This contrasts with what Charles (2003) finds for prime working-age US men. The difference is most likely attributable to the lack of incentives for DI recipients in continental Europe to move off the roll.

In Austria and the Netherlands, but not in Denmark, the impact on employment is greater for women. In the same two countries, the effects are greater on older persons and blue collar (Austria) or low-income (the Netherlands) workers. This is consistent with evidence from the UK showing that older poor individuals are more likely exit the labor force by entering disability insurance (Banks, 2006). It is also consistent with

[22] There is no significant impact of a road accident on the employment probability of women.

US evidence from the PSID that earnings losses from ill-health are larger and more sustained for nonwhites and the poorly educated (Charles, 2003), characteristics associated with working in industries and occupations in which productivity is more contingent on physical health and that identify low-skilled workers for whom, as observed in Section 17.3.3.2, labor market opportunities have deteriorated and dependence on DI has increased. The less privileged may not only be more likely to be struck by ill-health, but their employment and incomes are also more contingent on their health. Ill-health may increase economic inequality through both its skewed incidence and its differential effect.

In the Netherlands, an acute hospital admission results in an average reduction of around 5% in personal income 2 years after the health shock, with little or no recovery over the following 4 years (García Gómez et al., 2013). For individuals who remain in employment, income falls by only 3%, indicative of very modest reductions in wages and hours of work at the intensive margin. Those moving onto DI experience an income loss of one-third, which is broadly consistent with the DI replacement rate. Although this is a substantial drop, the moral hazard effects of providing more complete income protection are likely to be large in a country where at one time 10% of the working-age population was on DI. But Denmark does offer even greater insurance, with an average 12% average drop in male earnings maintained for 6 years after a road accident offset by a rise in transfer income such that there is no significant change in total income (Moller Dano, 2005). This does not imply that there is complete insurance. Presumably, the 12% who lose employment do experience income losses, but this is not reported. For women, there is no significant drop in earnings or loss of income. These findings are consistent with the income consequences of ill-health being muted in a country, such as Denmark, with a generous welfare state. But one should keep in mind that it is the effect of a road accident that is estimated. According to McClellan (1998), accidents should have the mildest economic consequences because the induced health change may be temporary, and longer term functioning and survival expectations may be little affected. This is the limitation of the described approach. Although one can be confident that the estimate does accurately capture the effect of the health change studied, the rarity of the narrowly defined event reduces the relevance of the evidence to the broader question of the extent to which the overall variation in health contributes to observed economic inequality in the population.

17.3.4 Early-Life Health Determinants of Later-Life Economic Inequality

The focus on the income effects of ill-health in adulthood until this point risks missing much of the action. A rapidly growing literature, to which Janet Currie and James Heckman are leading contributors, argues that early-life—even prebirth—and childhood conditions, including health, explain much of the variation in economic outcomes across adults (Almond and Currie, 2012; Cunha et al., 2006; Currie, 2009; Heckman, 2007; Heckman et al., 2006). It has been estimated that a staggering 50% of inequality in the present value of lifetime earnings in the USA can be explained by factors known at age 18 (Cunha and Heckman, 2009). Although most, such as parental occupation, are not directly health-related, exposure to health risks in the womb, infancy, and childhood is a potentially important component of these economically significant conditions. Ill-health in early life may directly constrain health capital in later life and impede the accumulation of nonhealth human capital. Both effects would reduce earnings potential.

One may distinguish three broad mechanisms through which early-life health may impact on economic outcomes in adulthood, differentiated by the life stage in which they become manifest. Nutritional deprivation and exposure to health risks *in utero* and in infancy can directly impair cognitive functioning and lead to childhood health problems that interfere with the acquisition of cognitive, and possibly noncognitive, skills. A second route is through education. Ill-health in childhood and adolescence may restrict opportunities to acquire education, and impaired cognitive functioning arising from insults to health in infancy may reduce the efficiency of schooling in producing educational qualifications. The third mechanism operates through health capital, as opposed to other forms of human capital. Ill-health may persist from childhood to adulthood. More dramatically, exposure to health risks in the womb may do lasting physiological damage, which becomes manifest with the onset of disease in middle age. Earnings may subsequently fall, as is clear from the evidence reviewed in the previous section.

In the following subsections, we discuss each of these three broad mechanisms by which infant and childhood health may constrain economic success in adulthood. This takes us into territory that has traditionally been the domain of psychology and epidemiology but in which economists are increasingly daring to venture. We end the section by summarizing and evaluating the evidence on the extent to which adult economic outcomes are determined by early-life health conditions. We do not provide detailed reviews of the rapidly growing literatures, which have already been provided by researchers with far greater expertise (Almond and Currie, 2011, 2012; Cunha et al., 2006; Currie, 2009). Our focus is on what the literature has to say about the contribution of health to income inequality and to understanding the association between income and health in adulthood.

17.3.4.1 Health, Cognitive, and Noncognitive Capabilities

There is abundant evidence that cognitive functioning is a strong predictor of wages (Cawley et al., 2001; Herrnstein and Murray, 1994; Jencks, 1979). We are interested in whether cognitive functioning measured during childhood determines economic success later in life and whether childhood cognitive function is in part determined by health in infancy. Case and Paxson (2008) provide indirect evidence of the first relationship. Using longitudinal data from the UK, they show the strong positive correlation between earnings and adult height.[23] This relationship is well established, but it falls greatly in magnitude and becomes insignificant when cognitive functioning in childhood, which is shown to be a strong predictor of wages, is controlled for. The change resulting from the cognition control is consistent with early-life nutrition producing both cognition and height and only the former impacting on earnings. After dealing with bias arising from cognition measured in adolescence being a product of (endogenously chosen) schooling, Heckman et al. (2006) find that cognitive functioning is an important determinant of schooling, employment, occupation, and wages.

So, cognitive functioning in childhood has economic consequences. Do we know that cognition itself is contingent on health in infancy? Through animal experiments, neuroscience has identified the biological and neurological processes that link undernutrition, as well as nutrient deficiency and exposure to toxins, *in utero* and infancy to impeded development of the brain (Grantham-McGregor et al., 2007). Birth weight, which is an indicator of exposure to health risks *in utero*, particularly nutritional deprivation, is the most frequently used indicator of the health of humans at birth. Epidemiological studies confirm that low birth weight is associated with low IQ (Breslau et al., 1994), along with a host of other child health problems including asthma (Nepomnyaschy and Reichman, 2006), behavioral problems including attention deficit hyperactivity disorder (ADHD) (Hayes and Sharif, 2009; Loe et al., 2011), slower motor and social development (Hediger et al., 2002), and depression (Costello et al., 2007). While consistent with the neuroscience, one cannot read too much into a simple correlation because low birth weight could reflect the behavior of the pregnant mother correlated with later investments in the child that influence cognitive functioning. Variation in birth weight within siblings and twins has been used to reduce the risk of such bias. This approach has produced evidence that the smaller sibling or twin tends to have a lower IQ at age 7 in Scotland (Lawlor et al., 2006) and on entrance to the military in Norway (Black et al., 2007). The latter finding is more difficult to interpret because IQ in young adulthood could reflect differential investment in education.

In low- and middle-income countries, where the nutritional and micronutrient deficiencies are obviously much more pronounced, there is clearer evidence from randomly assigned nutrition supplementation programs that better nutrition improves cognitive functioning and raises educational attainment (Grantham-McGregor et al., 1991, 2007; Pollitt et al., 1993; Walker et al., 2005). Currie (2009) cites evidence showing that even in the US children of mothers included (not randomly) in a nutritional program during pregnancy achieve higher test scores.

Heckman and his collaborators present evidence demonstrating that noncognitive skills developed in childhood are as important, possibly even more important, than

[23] In the USA and the UK, someone who is 1 in. taller has, on average, 1.5–2% higher earnings (Case and Paxson, 2008).

cognitive functioning in explaining economic outcomes in adulthood (Cunha and Heckman, 2009; Heckman, 2007; Heckman et al., 2006). Noncognitive capabilities refer to personality traits, such as self-esteem, perseverance, dependency, consistency, patience, and optimism, which may be considered to be determinants or aspects of preferences over risk and the timing of consumption. Currie (2009) points out that some noncognitive skills are closely related to, or are highly contingent on, mental health conditions. She cites a number of studies presenting evidence that child behavioral problems, such as ADHD and aggression, are strong predictors of lower cognitive functioning, educational attainment, and economic outcomes. Most of the studies control only for observables, but Currie and Stabile (2006) use sibling fixed effects and find that children in both the USA and Canada with high ADHD scores at younger ages had lower cognitive functioning (math and reading test scores) at age 11 and were more likely to be admitted into special education and to have repeated a grade. Given that behavioral mental health problems are so prevalent in children, they are potentially an important part of the link between child health and adult economic circumstances.

Although the evidence base does need strengthening, we believe that there are sufficient conceptual grounds for expecting health in infancy and early childhood to emerge as an important constraint on the formation of cognitive and noncognitive skills that are increasingly recognized as important determinants of labor market success. Within the framework developed by Cunha and Heckman (2008), the importance of these skills in the generation and reproduction of economic inequality derives from the fact that they are malleable. Skills are produced through parental investments, the level and efficiency of which are likely to depend on the socioeconomic environment of the child's family and neighborhood. Heckman (2007) hypothesizes that there may be complementarities between investments in health and cognition. Bad luck of the draw that leaves a child with a deficit of one may make it more difficult to raise the other through investments. An early-life health shock could leave a child frail, with limited capacity to respond to the stimuli that can raise cognitive functioning. Facing a higher price for a marginal gain in functioning, parents may invest less in the development of the child's skills. Thus, sickly kids may be doubly penalized with a deficit in both health and human capital.

17.3.4.2 Education

Education is a potential conduit that links income to health not only over the life cycle but also across generations. Health problems in childhood may directly constrain the acquisition of education, as well as weaken incentives for investment in schooling, with long-run consequences for income. If the children of poorer and less healthy parents are more likely to experience illness, then the interference of health capital in the acquisition of other forms of human capital could contribute to the intergenerational transmission of income (Currie, 2009). Health determination of income distribution may operate with a very long lag.

According to life course epidemiology (Kuh and Ben-Shlomo, 1997; Kuh and Wadsworth, 1993; Wadsworth and Kuh, 1997), childhood illness, which may arise from social deprivation, not only has a permanent effect on health, but it also interferes with education. As a result, occupational opportunities are limited in young adulthood, which may further reduce adult health, and lifetime earnings potential is constrained. Health and income in adulthood are correlated because they are both determined by childhood illness.

The most straightforward way in which health may impact education is through the interruption of schooling. In low-income countries, this could be an important constraint. In high-income countries, it seems less relevant. Currie (2009), citing Grossman and Kaestner (1997), notes that differences in school absence rates between healthy and unhealthy US kids are too small to lead to a strong correlation between health and educational attainment. Any impact of child ill-health on knowledge and skills acquisition more likely operates through the channel examined in the last subsection constrained cognitive functioning and impaired efficiency of learning.

Using a cohort of Britons born in 1958 and controlling for childhood socioeconomic status (SES), Case et al. (2005) find that a health problem in childhood is correlated with lower educational attainment. But poor health and low education could both result from parents who make little investment in the human capital of their offspring. A partial solution is to go one stage back in the child's development when parental behavior can exercise less influence and examine the association between birth weight and educational attainment. We noted above the evidence of the impact of birth weight on cognition. If this is the channel, we now need to establish the next link in the chain, to educational outcomes. There are many epidemiological and social science studies showing that children with very low, or even low, birth weight tend to perform poorly in school (Case et al., 2005; Currie and Hyson, 1999; Hille et al., 1994; Kirkegaard et al., 2006; Saigal et al., 1991). The causality of the relationship is given credence by sibling and twin difference studies of sufficiently large samples from high-income countries across three continents. These studies show lower educational attainment by the children who were smaller at birth (Behrman and Rosenzweig, 2004; Black et al., 2007; Johnson and Schoeni, 2011; Lin and Liu, 2009; Oreopoulos et al., 2008; Royer, 2009). There is, however, variation in the magnitudes of the estimates. For example, using US PSID data, Johnson and Schoeni (2011) find that low birth weight increases the probability of dropping out of high school by one-third, with part of the effect appearing to operate through impaired cognitive functioning. Royer (2009), using data on twins who both became mothers in California, finds a rather small average effect.[24] A claimed feasible 250 g increase in birth weight would raise schooling by only 0.04 of a year.

[24] A priori, restriction of the sample to twin pairs who both became mothers risks selection bias, although analysis conducted by the author suggests that this is not large.

A couple of US studies only find significant or substantial effects for children born into poor families or neighborhoods (Conley and Bennett, 2001; Currie and Moretti, 2007). This is consistent with wealthier parents being able to compensate for a health disadvantage in early life, when poorer parents lack the means to invest in medical care or other health and educational inputs. Other studies conducted with British (Currie and Hyson, 1999), Norwegian (Black et al., 2007), and Canadian (Oreopoulos et al., 2008) data find no evidence of this heterogeneity. Hasty attribution of this discrepancy to the equalizing effect of universal health care coverage in Europe and Canada, but not in the USA, would be foolhardy. The 1958 British cohort would have had access to only rudimentary medical interventions for the treatment of low weight babies by today's standards (Almond and Currie, 2012). But it could be that little treatment was available for both rich and poor low-birth-weight babies in 1950s Britain. The pertinent question is whether the gradient observed in the more recent US data arises from differential access to effective medical care and possibly other corrective interventions. It may also be that the incentives parents have to make investments that compensate for poor child health differ between the rich and the poor. A particularly tight budget constraint may not allow investment in all offspring. It can then be optimal to concentrate investments on the child with the best chance at the expense of the child the parents observe to be frail and likely to struggle in life in any case (Almond and Currie, 2012).[25]

There is emerging evidence of a link between measures of school performance, or educational attainment, and *in utero* exposure to health risks through disease (Almond, 2006; Kelly, 2011), radiation (Almond et al., 2009), and maternal alcohol consumption (Nilsson, 2009). This is arguably a more convincing strategy to identify the effect of early-life health on education than that of twin differences because it uses variation in infant health risks that is external to the family environment and so is more plausibly exogenous to factors that also impact on schooling. Even more convincing is evidence that treatment of children for disease raises school attendance (Bleakley, 2007; Miguel and Kremer, 2004).

The evidence is convincing that health status at the time of birth impacts positively on educational attainment. How does this effect operate? Does fetal distress permanently damage cognitive functioning, which interferes with knowledge acquisition and skills development throughout school? Or do children who are frail at birth subsequently develop health conditions and illnesses at preschool ages that delay development and place the child at a disadvantage throughout his or her schooling? Or do health problems

[25] In general, the investment response will depend on the technology of human-capital production and the parental aversion to inequality across children (Almond and Currie, 2011). When technology is such that early-life health shocks can only be partially offset by later childhood investments, it can be optimal to reinforce the shock by withdrawing investment in the frail child irrespective of the degree of inequality aversion. This is more likely to occur as household income falls. See Appendix C and footnote 7 in Almond and Currie (2011).

in the preschool and early school years lead to additional health problems in later school life at ages when qualifications must be obtained? Currie et al. (2010) are able to address these types of questions by using rich Canadian data following children from birth into young adulthood. Controlling for sibling fixed effects, birth weight, and congenital and perinatal abnormalities, which always have effects, major physical health conditions and injuries in the preschool (0–3) and early school (4–8) years impact on educational attainment only because they raise the probability of experiencing similar conditions in later school years. The process is different with respect to mental health conditions. Behavioral problems at all ages, including the preschool years, directly reduce educational attainment. From this and other evidence, Almond and Currie (2012) conclude that there is an important effect of early-life health on human-capital acquisition and later-life outcomes. Exposure to adverse conditions *in utero* appears to exert a stronger effect than postnatal health, although mental health conditions identified in the preschool years have lasting effects.

Besides a *technological* effect of childhood ill-health on the acquisition of education, if the reduction in health is permanent, then there may be an *incentive* effect operating through life expectancy. A longer life may raise education by lengthening the period over which the return on this investment can be reaped (Ben-Porath, 1967). This further strengthens the extent to which dispersion in health implies dispersion in earnings potential. There is an inequality increasing concentration of health and human capital in the same individuals.

17.3.4.3 Fetal Origins Hypothesis

Early-life health potentially determines later-life economic outcomes not only via education and skills acquisition but more directly through health problems that interfere with productive work in adulthood. The *fetal origins hypothesis*, proposed by David Barker, is that chronic diseases, principally coronary heart diseases but also related diseases such as type 2 diabetes, originate in nutritional deprivation in gestation and infancy (Barker, 1992, 1995; Barker et al., 1993). Nutrition-induced stresses placed on the fetus at critical stages of development alter the physiology of vital organs, particularly the heart, which makes them susceptible to failure in middle and old age, inducing the onset of chronic disease.[26] Metabolism can also be detrimentally affected by *in utero* nutritional deprivation such that the risk of obesity is raised.

One cause of fetal stress that has been exploited to estimate long-term economic effects is *in utero* exposure to the 1918–1919 Spanish flu (Almond, 2006), which has been shown to increase the incidence of stroke, diabetes, and hearing, seeing, and mobility impairments (Almond and Mazumder, 2005), as well as cardiovascular disease

[26] The theory is not free of sceptics in medicine and epidemiology. See, for example, Paneth and Susser (1995) who call for clearer elucidation of precise physiological mechanisms and their testing.

(Mazumder et al., 2010). As is clear from the evidence reviewed in Section 17.3.3.5, chronic illnesses such as these have negative impacts on employment and earnings. Economic consequences of the *fetal origins hypothesis* are a topic that is increasingly receiving attention (Almond and Currie, 2011, 2012; Currie, 2009). We discuss some of the evidence in Section 17.3.4.4.

17.3.4.4 Economic Consequences of Early-Life Ill-Health

The previous three subsections establish that ill-health in early life and childhood constrains the acquisition of nonhealth human capital and directly impinges on adult health. With these disadvantages, one expects adults who were sick, or exposed to health risks as infants and children to be less well-off. What is the evidence that frailty of health at the beginning of life leads to economic disadvantage later in life?

Analyses of the 1958 British birth cohort establish that low birth weight is not only associated with lower educational attainment but also with lower employment, greater likelihood of being engaged in manual labor, and lower wages (Case et al., 2005; Currie and Hyson, 1999). Sibling and twin fixed effects studies produce evidence that lower birth weight reduces wages in Minnesota (Behrman and Rosenzweig, 2004) and reduces employment and earnings in the USA (Johnson and Schoeni, 2011).[27] It also reduces earnings (but not employment) in Norway (Black et al., 2007).[28] In addition, low birth weight increases welfare dependency in Canada (Currie et al., 2010; Oreopoulos et al., 2008) and may (Currie and Moretti, 2007) or may not (Royer, 2009) increase the likelihood of living in a poor neighborhood of California.

Johnson and Schoeni (2011) find that only around 10% of the effect of low birth weight on earnings operates through years of schooling.[29] The fact that most of the health effect does not operate via human capital is consistent with the findings of other studies (Luo and Waite, 2005; Persico et al., 2004; Smith, 2009). Using the 1970 British Cohort Study, Conti et al. (2010a,b) do not find any selection into postcompulsory schooling on the basis of child health for males and only a weak effect for females, although there is very strong selection on cognitive and noncognitive skills. There is a direct effect of child health on wages for males, but the weak wage effect for females does run via education.

[27] Low birth weight is estimated to reduce labor-force participation by 5% points and earnings (given employment) by around 15%.

[28] A 10% increase in birth weight is estimated to raise earnings of the full-time employed by 1%, an effect equivalent to that of about 3 months of education.

[29] In interpreting this finding, one should bear in mind that birth weight is estimated to have a small and insignificant impact on years of education. Rather, its stronger impacts are on the probability of high school graduation and on test scores. If these intermediate outcomes were considered, then, presumably, the proportion of the effect of birth weight on earnings that operates through human capital would be revealed to be greater.

The health measures employed by Conti et al. are height and head circumference at age 10. Although these measures should pick up nutritional deprivation at early ages, they will miss many other child health conditions. Estimating the long-term economic impact of child health, more generally defined, is made difficult by the scarcity of suitable longitudinal data and the challenge of measuring the general health status of children. Using the US PSID, Smith (2009) estimates that an adult retrospectively reporting *excellent* or *very good* health in childhood earns 24% more than his or her sibling who reports less than *good* health in childhood. There are no differences in education, so again most of the effect, if we are prepared to label the earnings difference as such, does not appear to operate through human capital.

In an intriguing paper, Almond (2006) finds dramatic effects of *in utero* exposure to the 1919 Spanish flu on economic outcomes. Earnings of male workers were reduced by 5–9%. Applying standard estimates of the return to education to the impact on schooling (length reduced by 5 months, on average, and high school graduation probability down by 13–15%), the indirect effect through education explains around one-half of the earnings effect. So, although there is support for the *pathways model* of life course epidemiology that proposes that the infant health effect operates (partly) through education (Hertzman et al., 2001; Kuh et al., 2003), this does not tell the full story. There appears to be a substantial direct effect of fetal conditions on earnings.

Welfare payments, which include DI, are higher for those exposed to the flu. There is also evidence of a substantial effect through occupation. For men, exposure results in a drop of around 6% on a hierarchical index of occupational status. Total income is reduced by about 6.4% for men, and the probability of being poor is increased by as much as 15% points.[30] Identification of these effects from prebirth conditions cuts through the endogeneity problems that plague the evidence on the economic consequences of ill-health reviewed in Section 17.3.3.5. The limitation is that the link between health and economic outcomes is less transparent. The identification requires a leap of faith that it is the Spanish flu and not some other peculiarity of the 1919 birth cohort that is responsible for the effects.[31]

[30] Poverty is defined as an income below 150% of the poverty line. The effects presented here for occupation and income are computed as the average of the effects for the 3 census years presented in Table 17.1, each of which is scaled by a one-third infection rate and expressed relative to the control group means in Table 17.2 in Almond (2006).

[31] Additional evidence of the long-term economic effects of early-life health conditions includes Nilsson's (2009) finding that increased *in utero* exposure to alcohol due to mothers drinking more following a liberalization of licensing laws in Sweden not only reduced years of schooling but also decreased earnings and increased welfare dependency. Based on the effects of a program to eradicate hookworm in the American South in the first quarter of the twentieth century, Bleakley (2007) estimates that infection with the disease in childhood reduced the probability of school enrollment by 20% and reduced wages by around 40%.

Subject to this caveat, this study has important implications for the interpretation of income distribution and its association with health. Sizable differences in incomes, and in its education and occupation determinants, appear to be attributable to differences in health conditions at the very beginning of life. Policies that can improve the early-life conditions of the most vulnerable infants can potentially compress the distribution of income decades later by ensuring that there are fewer physically frail individuals who can become the most economically deprived. Almond (2006) draws attention to the tremendous racial disparity in early-life conditions in the USA, where nonwhites are exposed to twice the infant mortality rate of whites. His estimates suggest that this health inequality may not only be a consequence of current economic inequality but also a potentially important cause of future economic inequality.

Overall, the evidence points to the health environment in which a child is conceived and delivered exerting a lifelong effect on economic opportunities. Currie (2009) notes that it is difficult to gauge the magnitude of the effect, given the variety of health risks and measures that have been employed by researchers and the differing study contexts. However, the evidence would seem sufficient to conclude that it is not a negligible effect. A further tentative conclusion is that the effect of health risks in infanthood on economic outcomes, via adult health, is stronger than the effect that operates via educational attainment. This is sometimes claimed as support for the *fetal origins hypothesis*. Strictly, it is not. The hypothesis claims a direct causal link between fetal health and the onset of chronic disease in middle age. Yet, a number of the studies purported to find evidence consistent with the hypothesis examine younger adults, and the health measures used often do not identify the cardiovascular conditions that are triggered by fetal stress according to the theory. Adult health can also be related to health conditions in childhood because illness is persistent and cumulative, which is a feature of the *accumulation model* of epidemiology (Kuh et al., 2003; Riley, 1989).

It is safe to conclude that health in early life is relevant to economic circumstances in adult life. Establishing precisely why and to what extent is a challenging research agenda to be tackled by both economists and epidemiologists.

17.3.5 Health and Occupation

The *life course model* identifies entry into the labor market as an important stage at which the relationship between health and SES is strengthened (Kuh and Wadsworth, 1993). Frail young adults face a narrower choice of entry-level jobs both because their education has been constrained by ill-health and because persistent health conditions directly impede productivity or provoke discrimination. The evidence reviewed in the previous section suggests that the direct productivity effect is stronger than the education effect. The income gradient in health may then partly reflect selection of the less healthy into lower-paid jobs, as well as sickness impeding movement up the career ladder.

This mechanism has been the focus of the epidemiological literature that has considered the extent to which causality runs from health to SES. The literature generally concludes that health-related occupation selection and social mobility are of insufficient magnitudes to make important contributions to the observed socioeconomic gradient in health (Chandola et al., 2003; Power et al., 1996, 1998). In this literature, SES is typically measured by occupation, social class, or employment grade. What is found is that, although job changes are related to health, movements of the more healthy into "better jobs" are insufficient to explain the observed health disparity across the occupational hierarchy (Chandola et al., 2003). The same need not follow with respect to the explanation of health differences across more economic dimensions of SES, such as earnings or income (Adda et al., 2003). As is hopefully clear by now, ill-health can impact on income through many channels other than occupation, and thus, even if this is not an important mechanism, it certainly does not follow that the income (rather than the occupation) gradient in health is mainly attributable to causality from income to health, rather than vice versa. Researchers from the economics discipline are generally more sympathetic to the selection hypothesis (Deaton, 2002, forthcoming; Smith, 2004). In particular, they recognize and emphasize the potential evolution of an individual's career and health based on his early-life health and other experiences, as discussed in the previous section (Case and Paxson, 2011).

The health-related selection of occupation need not necessarily steepen the income gradient in health. If productivity varies with health, and productivity is unobservable, then wages will vary with health only if the latter is observable (Strauss and Thomas, 1998). If health is not observable, then the healthy would be expected to sort into occupations in which productivity is less difficult to observe. Particularly in low-income countries, the healthy may stick to self-employment in which there is no issue of productivity verification and earnings are not set at some average over higher and lower levels of productivity. But such occupations may offer lower earnings potential, perhaps because there is less capital per unit of labor. Wage differences by health would be more compressed than in a situation in which there was no sorting.

The evidence reviewed in Section 17.3.2.4 generally provides estimates of the impact of health on the wage rate conditional on occupation. But there could be an additional effect on earnings through occupation itself. The evidence on job change after the onset of a health condition is relatively sparse. Now rather old data reveal that around one quarter of males and one-fifth of females in the USA with a work-limiting health problem report having moved to jobs more compatible with their conditions (Daly and Bound, 1996). Older workers and high school dropouts are less likely to change jobs. Charles (2003) argues that older workers are less likely to retrain for the purpose of accommodating a disability, because they have less time remaining to reap a return on this investment. He finds evidence consistent with this hypothesis in the US PSID. The low-educated may be less likely to adapt because they lack the general human capital that raises the efficiency of specific investments. It is more difficult to move from manual to nonmanual labor if basic reading and writing skills are lacking. This may partly explain why lower-educated, lower-skilled, and lower-paid workers are more likely to exit the labor force when struck down by a disability (Banks, 2006; García Gómez et al., 2013; Halla and Zweimuller, 2013) and, in the USA, are less likely to recover earnings lost after the initial onset of a disability (Charles, 2003).

17.3.6 Health and Household Income

The impact of health on the distribution of household incomes may differ from that on the distribution of individual incomes for two principal reasons. First, health may affect the formation and dissolution of households. Second, the illness of one household member may provoke a response from the labor supply of others.

Poor health may make it more difficult to find a partner. Limitations in functioning, caring needs, and reduced expected longevity may make a disabled or chronically ill person a less attractive proposition in what some economists refer to as the *marriage market*. On top of the direct effect, there may be a reinforcing indirect effect operating through the impact of health on human-capital accumulation and earnings potential; marrying a sicker person, on average, means marrying a poorer person. The marriage vows of "in sickness and in health" appear to recognize the threat that illness poses to marriage. If less healthy people do have fewer opportunities to find and keep a partner, then one would expect that there will be health-related sorting. Then lower than average earnings of a disabled or sickly person will be compounded, not compensated, by the earnings of his or her spouse.

Although the idea that health influences marriage prospects has been around for some time (Carter and Glick, 1976; Sheps, 1961), there is little convincing evidence with which to judge its empirical validity. In part, this is because of the difficulty of separating healthy selection into marriage from the potentially beneficial effect of marriage on health (Goldman, 1993). Longitudinal data are required. Fu and Goldman (1994) find little evidence that health predicts the marriage behavior of young American adults. For US women there is evidence of health selection, however. Among young women not in full-time employment, those in better health are more likely to marry and less likely to break up (Waldron et al., 1996), and in a Californian sample of siblings, the sister with the lower birth weight is 3% points less likely to be married when she gives birth (Currie and Moretti, 2007).

Recognition of potential health gains from marriage introduces the possibility that there is negative health selection into the institution. The less healthy have more to gain from marriage. Lillard and Panis (1996) find evidence of such adverse selection among US men; the less healthy (re)marry sooner and remain married for longer. However, there is also selection on unobservables correlated with good health, and this dominates so that married men are healthier than their unmarried counterparts.

Within a household, ill-health of one partner could provoke two conflicting motivations for labor market activity of the other partner. On the one hand, reduced earnings of the disabled partner will generate an income effect that will motivate the spouse to replace those lost earnings through increased work effort. This is the added worker effect that is familiar in the unemployment literature. On the other hand, the disabled person's productivity may be reduced not only in the labor market, but also within the household. Limitations in functioning may reduce capacity to wash, dress, and feed oneself. Meeting these caring needs will place demands on the spouse's time. *A priori*, one cannot say which effect will dominate.

There appears to be a gender difference in the relative magnitude of the two effects, but the direction of this bias is not always consistent. Most US evidence finds that women are more likely to participate in employment when their husbands fall ill, but, if anything, male spouses are less likely to participate (Berger, 1983; Berger and Fleisher, 1984; Charles, 1999; Van Houtven and Coe, 2010). Coile (2004) finds no effect on the female spouse and only a small increase in the employment of men whose wives fall ill, however. The employment response of the spouse has also been found to depend on the type of health condition and the initial labor supply of the spouse (Blau and Riphahn, 1999; Siegel, 2006).

There is evidence from Germany (Riphan, 1999), Spain (García Gómez and Lopez-Nicolas, 2006), and the Netherlands (García Gómez et al., 2013) that ill-health reduces household income by more than the fall in the personal income of the person experiencing the health shock. For example, the Dutch study finds that an acute hospital admission reduces household income by 50% more than the reduction in income of the person admitted to hospital.

17.3.7 Health and Wealth

The impact of health on economic inequality may go beyond wage and income distributions to wealth distribution.[32] If ill-health reduces income through one or more of the mechanisms identified in the previous subsections, then opportunities to accumulate wealth over a lifetime will be constrained. Because the effect is accumulated, permanent differences in health will create greater variance in wealth than in income. In addition, ill-health may force depletion of wealth to pay for medical or nursing care. Less obviously, health may affect wealth through life expectancy and consequent saving incentives. The *horizon* and *uncertainty* effects of increased longevity both raise saving, while they have contradictory effects on labor supply (see Section 17.3.3.4). Those expecting to live

[32] Identification of the distribution of wealth itself can require taking an account of the distribution of health, or rather mortality. When one only observes the wealth of the deceased in the form of inheritances, then mortality multipliers need to be applied in order to infer the distribution of wealth among the living. Differential mortality may be taken into account (Atkinson and Harrison, 1978).

for longer will accumulate more wealth both to provide for an extended old age and because they face a lower risk of dying before having the opportunity to enjoy their savings.

Consistent with these mechanisms, the relationship between health and wealth is particularly strong. PSID data reveal that the median wealth in 1994 of a household whose head was in *excellent* health 10 years earlier was 268% greater than the median wealth of a household whose head had been in *poor* health (Smith, 1999). This wealth inequality grew both with the lapse of time since the difference in heath was recorded and with age, consistent with differential rates of wealth accumulation by health. A number of US studies have examined whether the strong positive relationship between health and wealth is due to causality from health to wealth, causality from wealth to health, or simply a spurious correlation.

The much-cited paper by Adams et al. (2003) analyzes panel data on a sample of the US population aged 70+ (Asset and Health Dynamics among the Oldest Old—AHEAD). Their focus on the elderly eliminates differential earnings as a mechanism through which health may contribute to differences in wealth accumulation. The null of no causation from both current and previous health, indicated by 19 conditions and SAH, to the change in wealth is rejected.[33] The authors' rigorous analysis leads them to emphasize that, although the result is consistent with a causal effect from health to wealth, they cannot rule out the possibility that it reflects model misspecification and/or time-invariant unobservable factors driving the evolution of both health and wealth.

Michaud and van Soest (2008) overcome both limitations and provide even more conclusive evidence of a causal effect from health to wealth. They use the HRS and allow for causality operating contemporaneously and with lags in both directions, along with unobservable heterogeneity. Health is measured by an index constructed from principal components analysis of SAH, major and minor conditions, ADL, depression score, and body mass index. The health of both the husband and the wife are found to impact on household wealth. The effect of the wife's health is immediate, but that of the husband's health is delayed.[34] This assumption is due to a gender difference in the type of ill-health that impacts on wealth. For both sexes there is a delayed impact of physical ill-health. But only the mental health of females has an impact on wealth that is immediate. Evidence of the causal impact of health on wealth is stronger for households that lack health insurance

[33] The null is rejected for total, liquid, and nonliquid wealth for couples in which both partners survive and single households (except nonliquid wealth). The null is not rejected in a number of cases for couples experiencing a death (see Adams et al., 2003, Table 11).

[34] Instruments are required in models that allow contemporaneous effects. The onset of major health conditions (cancer, heart condition, lung disease, and stroke) is used under the assumption that these critical illnesses only impact wealth though health and are not contemporaneously affected by changes in wealth. Essentially, this is the same identification assumption used by Smith (2004) and Wu (2003).

coverage, particularly those in which the wife succumbs to mental ill-health, suggesting that depletion of assets to pay for medical care is an important part of the effect.

Although the Michaud and van Soest study gives us good reason to believe that there is an effect of health on wealth, at least in the older US population, its use of a health index has the disadvantage of not producing an estimate of easily interpretable magnitude. Without allowing for unobservables that condition the evolution of both health and wealth, Smith (2004) estimates, using HRS data, that the income loss, medical expenses, and consequent forgoing of interest arising from the onset of a major health condition (see Section 17.3.3.5) accumulate over 8 years to an average loss of wealth of almost \$50,000. Most of this lost wealth is due to reduced earnings. Consequently, the wealth loss is considerably lower (\$11,350) for the older AHEAD cohort analyzed by Adams et al. The wealth loss resulting from the onset of a minor health condition is also much smaller (\$11,500). Recognizing that one-fifth of Americans aged 50+ experience the onset of a major health condition over an 8-year period and a further 30% incur a minor condition, Smith argues that the consequent wealth losses represent substantial effects of health on the distribution of wealth. Consistent with Michaud and van Soest (2008), the magnitude of the effect that does not operate though earnings losses is larger when the wife experiences the illness (Wu, 2003).[35] This effect relationship is explained by assets being run down to pay for general living expenses when the wife is no longer fit to perform household chores. Accordingly, it is not observed when the husband's health deteriorates.

Overall, the evidence is convincing that health constrains the accumulation of wealth, and illness speeds its depletion. The magnitude of the effect is likely to differ with the nature of the health condition and the means of financing both pensions and medical care.[36] In the USA, for which most evidence is available, the effect seems substantial. Variation in both health levels and rates of health depreciation with age may make substantial contributions to interhousehold inequality in wealth holdings.

17.3.8 Summary

Understanding the effects of health on income and wealth is important for the explanation of distributions of income and wealth and the interpretation of the economic

[35] This evidence comes from analysis of only the first two waves of the HRS and the contemporaneous relationship between changes in health and changes in wealth.

[36] Hurd and Kapteyn (2003) propose and confirm that the relationship between wealth and health at older ages is weaker in a country such as the Netherlands, where a greater share of retirement income is obtained from annuities, than in a country such as the USA, where savings and assets are more important sources of financing consumption in old age. Differential financing of medical care, and not only the source of retirement incomes, may contribute to this result. One expects a stronger correlation between wealth and health in countries, such as the USA, with less comprehensive public health insurance and where, until the 2010 Affordable Care Act, private insurance premiums could be related to pre-existing conditions.

gradient in health. We have identified a number of pathways through which health potentially impacts income and wealth, and the evidence suggests that many of these are empirically important. Ill-health can lead to a fall in wages at the margin, but wages are more likely to drop through job changes. In high-income economies, policies that constrain wage flexibility, such as effective minimum wage laws and antidiscrimination legislation, which has been strengthened considerably in relation to disability in Europe and the USA from the 1990s onward, can limit the downward pressure on wages, but this effect often comes at the cost of reducing employment opportunities for disabled individuals. Legislation frequently obligates employers to adapt workplace and employment conditions to accommodate disabled workers, and the responses of employers may vary across sectors and occupations. Changes toward a more disability-friendly workplace are evident to those of us working in white collar and professional occupations. Employers faced with difficulties in recruiting skilled workers may offer an accommodating work environment and flexible hours to secure a competitive advantage. In lower-paid sectors, employment decisions have become more short-term, however, and firms may be even less willing to invest in accommodating low-skilled workers with health problems. Although we are not aware of any evidence of such a heterogeneous employer response, it is a hypothesis worthy of investigation, not least because it implies that increasing general labor market inequality would generate an even greater increase in inequality between disabled workers with different skill sets.

Policy further conditions the labor market response to ill-health through DI, which protects incomes but increases the impact on employment through a sizable incentive effect. From the perspective of public finances, the moral hazard of DI is a legitimate and substantial concern. However, the increased responsiveness of employment to ill-health does not necessarily imply welfare loss. For individuals suffering health conditions that make work extremely uncomfortable, DI may, indeed should, allow these workers to withdraw from the labor force. In lower-income economies with less social protection, ill-health may have a smaller impact on employment, and perhaps even on money income, but it could contribute more to inequality in well-being, because the necessity to continue working with a disability reduces worker utility, both immediately through the discomfort of work and in the long term through reduced health. This point serves as a reminder that, although the focus of this chapter is on the relationship between two central dimensions of well-being, income and health, we are ultimately interested in the implications of this relationship for the distribution of well-being itself.

The impact of ill-health on household income can be substantially larger than that on the earnings of the disabled person, because of a spillover effect on the labor supply of the spouse. In addition, illness may reduce the likelihood of forming and maintaining marriage partnerships, although the evidence on this point is mixed. Through these two mechanisms, the contribution of ill-health to inequality in household incomes need not be less and may even be greater than its contribution to inequality in personal incomes. Beyond DI, policy can attenuate the contribution of ill-health to inequality in household disposable incomes through the tax and benefit system. High marginal tax rates, resulting from tax credits and other income-tested transfers, will moderate any increased inequality in gross household earnings arising from the distribution of health.

Perhaps the most important conclusion emerging from the literature, and emphasized by others (Almond and Currie, 2012; Currie, 2009; Heckman, 2007), is that ill-health can have a very long reach from exposure to health risks in childhood to constrained economic opportunities in adulthood. By constraining human-capital acquisition through education and skills formation, as well as persistent and delayed effects on health in adulthood, early-life and childhood health experiences can be important determinants of both income distribution and the observed income gradient in adult health.

A good deal of evidence supports the mechanisms through which health may impact income and wealth, but inferring the magnitude of each effect is far from easy. This is because, even with respect to a particular mechanism, there is not a single effect but many. Health is not unidimensional. Different dimensions of health will impact income and wealth through different routes and to different degrees. Using a general measure of health, such as self-assessed health, may provide some average effect over different types of health problems, but the usefulness of this average is questionable. Relatedly, measurement error in health variables has been a substantial obstacle to obtaining credible estimates of health on labor market outcomes. But this problem is receding. Longitudinal surveys, such as the HRS and its equivalents, increasingly contain detailed measures of specific health conditions, allowing researchers to exploit the timing of illness onset in order to identify the economic consequences of intrinsically interesting changes in health with precise medical meanings. Also promising is the increasing access of researchers to linked administrative registers on hospital admissions, social insurance, and tax files, which drastically reduce measurement error and provide very large samples from which the effects of specific medical conditions can be identified.

Ill-health reduces income and wealth. The contribution of health to economic inequality depends upon how it is distributed. If health variation is random, then it adds to the dispersion of income (wealth). In this case, the additional economic inequality may not be considered socially objectionable. Losing income as a consequence of illness may be seen as unlucky but not unjust, and insurance may be called for on grounds of efficiency. The consequences for economic inequality—both the nature of the impact and its normative interpretation—are quite different if ill-health is not distributed by the roll of the dice, however. The next section examines whether income and wealth exert causal effects on health. Irrespective of whether such effects exist, if individuals with lower potential incomes are more likely to fall sick, then the income distribution is skewed even more to the disadvantage of the poor. For example, assume that low education both reduces income and increases the likelihood of sickness. The poor are more likely to get sick, and because they are sick, they become even poorer. The income distribution gets stretched by the unequal incidence of illness and its impact on income. Economic inequality is generated by biases that place the socially underprivileged at a health disadvantage and then impose an economic penalty for this status. Thus, inequality is more likely to be considered morally objectionable than it would be if it arose from even-handed luck in the distribution of illness.

In addition to the uneven incidence of illness, the effect of health on economic inequality is likely exacerbated by heterogeneity in the impact of health on income across the income distribution. The employment and earnings of low-skilled, low-paid workers is more contingent on their health than is the case for higher-paid professionals. Not only do the socially disadvantaged face a higher incidence of illness, but they are also more economically vulnerable to it. DI provides a safety net, but its disincentive effects are stronger for the low-paid, and they are more responsive to these incentives, because labor market opportunities have been deteriorating for this group since the 1980s. As a result, the loss of employment following ill-health is both a contributor to economic inequality and a consequence of it.

We conjecture that ill-health contributes to economic inequality not merely by adding noise to the distribution of income (wealth) but by further reducing the incomes (wealth) of those who would be located toward the bottom of the distribution in any case. This could occur even without low income or wealth reducing health. We now turn to the question of whether there is a causal effect from the economic to the health domain of well-being.

17.4. ECONOMIC DETERMINATION OF HEALTH INEQUALITY

17.4.1 Overview

If, as one would expect, health is a normal good, then the financially better-off will demand more of it. Whether this inflated demand is realized will depend on how health-enhancing and health-depleting goods are allocated. If medical care is delivered through the market, then the rich will be both willing and able to afford more effective treatment when illness strikes. But few countries, particularly high-income ones, leave the distribution of health care entirely to the market. Public provision of care to the poor and elderly, or even universal provision of care to the entire population, should constrain health differences that arise from variation in the individual's willingness and ability to pay for medicine. But other goods that are beneficial to health, such as quality housing, safe neighborhoods, and education, are at least partially allocated by the market and provide an opportunity for income to "buy" health. The direction of the relationship is not, however, unambiguous. Whether inequality in health reflects economic inequality will depend on the extent to which the greater demand of richer individuals for health is offset by their higher demand for the pleasures of alcohol, smoking, and rich foods that higher income makes affordable.

Whether the better-off should be expected to be in better health will also depend on the source of their economic advantage. If it arises from higher earnings potential, then the health effect is ambiguous. This insight emerges from Grossman's (1972a) health-capital model according to which health is demanded for a direct utility benefit—feeling sick is uncomfortable—and a production benefit—less time is lost to sickness and so more is available for work. Health is produced by investment in medical care, exercise, healthy eating, and so on. In the pure investment version of the model, which incorporates the production benefit only, a wage increase has two conflicting effects on health. A higher wage implies a higher value for a given increase in productive time, which would lead higher-waged individuals to invest more in health. But the marginal cost of the time input into health investment also increases. More earnings are lost visiting the doctor, jogging, and so on. The net effect is positive, provided that market goods, such as medical care and nutritious food, are used in the production of health in addition to the individual's own time input (Grossman, 2000). But in the pure consumption version of the model, which confines attention to the direct utility benefit of health, the fact that the time cost of producing health is less than the total cost is not sufficient to create a positive substitution effect from a wage increase. The relative intensity with which time is used in the production of health must be less than the relative time input into the production of other commodities that generate utility. Otherwise, a wage increase implies a rise in the relative price of health. Maintaining one's health is likely to be more time-intensive than many other activities that generate sources of utility, and thus, a negative pure wage effect certainly cannot be ruled out.

Empirically determining the extent to which economic advantage bestows health advantage is complicated greatly by the multitude of mechanisms, identified in the previous section, through which health impacts economic circumstances, as well as the plethora of *unobservables*, such as risk attitudes, time preferences, and genetics, that can influence investments in health and other human and financial capital. Fixed effects methods deal with the latter problem but are powerless against the former simultaneity problem, and early attempts to tackle both problems tended to rely on instruments of questionable validity.[37] Since the beginning of the twenty-first century, researchers have increasingly studied changes in the health response due to more plausibly exogenous sources of variation in income or wealth, such as sudden policy reforms, stock market volatility, or windfall gains. The weakness of this strategy is that it employs a form of economic variation that, while exogenous, does not correspond to the variation that can plausibly impact health. The problem is exacerbated by the fact that health does not respond immediately to a change in demand. Even if a windfall gain from a stock market

[37] For example, Ettner (1996) instruments the wage rate with work experience and state unemployment rate, and unearned income with parental and spousal education. Work experience may be correlated with the evolution of health, and parental education could have a direct effect on health.

boom, inheritance, or lottery win does induce a rise in an individual's desired level of health, achieving this improved health will take years of investment in preventive medical care, diet, and so on. Even long panels may be insufficient to observe this process. There is a greater chance of identifying the income effect on health determinants than on health itself.

Many chronic health problems are also unlikely to be provoked by sudden changes in income or wealth, but they may, in part, result from long-term exposure to unhealthy living conditions experienced by the poor. The time lag in such an effect, along with the obvious endogeneity issues, make its identification challenging. Most of the existing evidence does not relate to such long-term relationships, and this lack of applicable data must be kept in mind when interpreting the evidence from shorter-term variation that tends to show no, or a weak, effect of income or wealth on health, at least in higher-income countries.

In this section, we begin our review of the evidence by examining the impact of income and wealth on health in adulthood. Most of this evidence comes from high-income countries. We then turn to mechanisms and look at the evidence that economic resources impact health behavior and utilization of medical care. The penultimate subsection examines the evidence for an impact of household economic circumstances on child health. Much of this evidence has been collected from low- and middle-income countries, or it refers to low-income populations in high-income countries.[38]

17.4.2 Income and Wealth Effects on Adult Health

17.4.2.1 Causality Tests

Perhaps the most influential examination of the economic determinants of health conducted since the turn of the century is the study by Adams et al. (2003). Adams and colleagues recognize the difficulty of finding plausibly exogenous instruments for economic ircumstances that provide variation relevant to mechanisms of causation. So, they concentrate on the less demanding task of testing for the absence of causal income and wealth effects (and other dimensions of SES) on health among elderly (70+) Americans (see Table 17.5 for details of this study and other evidence relating to the USA). Their focus on an elderly sample neatly sidesteps the reverse causality from health to earnings that would likely occur in a sample of working-age individuals, and this complication is further avoided by adopting the concept of Granger (1969) causality and testing whether, conditional on lagged health, current health is uncorrelated with lagged income (wealth). This involves imposition of an assumption that there is no contemporaneous impact of income (or wealth) on health. The authors argue this plausible and strive to weaken it

[38] We purposefully do not cover macro studies that attempt to identify the impact of country GDP on the mortality rate (see Pritchett and Summers, 1996), because such studies tell us nothing about whether income is a determinant of health within any country.

Table 17.5 US evidence of income and wealth effects on adult health and health behavior

Authors	Data	Sample	Income/ wealth measure[a]	Measures of health/ health behavior	Estimator	IV for income/ wealth	Effect on health	Effect on health behavior or medical care use	Remarks
Carman (2013)	PSID 1984–2007	Adults	Inheritance	SAH	FE ordered logit		Overall—none Males—negative		Positive effect of anticipated inheritance
McInerney et al. (2013)	HRS 2006 and 2008	Adults 50+	Drop in nonhousing wealth	CES-D depression score, anxiety, medication, SAH	IV in 1st differences	Interview date post 2008 stock market crash	↑ CES-D ↑ 1.4 ppt prob. depressed ↓ prob. of *good* health	↑ prob. antidepressant use for large wealth losses	
Schwandt (2013)	HRS 1992–2008	Retirees reporting wealth, income and stocks	Change in stock value as % lifetime wealth	Onset health conditions (hypertension, heart disease, stroke, etc.), SAH, CES-D depression score, survival	OLS on change in health (2SLS in robustness checks)	Identification of change in S&P 500 index	↓ onset health conditions ↑ reported change in SAH ↓ CES-D, ↑ survival ↓ hypertension, heart disease, stroke, psychiatric problems No signif. effect on diabetes, cancer, lung disease, arthritis		
Van Kippersluis and Galama (2013)	HRS 1992–2010	50+	Household wealth	Smoking and drinking	IV FE	Inheritances		↑ moderate, not excessive, drinking. ↑ smoking	
Kim and Ruhm (2012)	HRS 1992–2006	50–60 at baseline	Inheritances (above and below $10,000)	Mortality, SAH, ADL, CES-D	Discrete time hazard (logit), LPM	Inheritance also used as IV for household income	Mortality—no effect Health—no robust evidence of effects		Effects on medical care and behavior explored
Stowasser et al. (2012)	HRS and AHEAD 1992–2008	50+	Liquid and nonliquid wealth income	As Adams et al. (2003)	As Adams et al. (2003)		Noncausality rejected for many conditions		Rejection of noncausality does not→causality

Table 17.5 US evidence of income and wealth effects on adult health and health behavior—cont'd

Authors	Data	Sample	Income/ wealth measure	Measures of health/ health behavior	Estimator	IV for income/ wealth	Effect on health	Effect on health behavior or medical care use	Remarks
Goda et al. (2011)	AHEAD 1993–1995	70+ at baseline	Social Security (SS) income	Use of home care or nursing home	IV probit	SS notch		↑ home care, ↓ nursing home care	
Salm (2011)	US Vital statistics 1900–1917	Union Army Veterans	Pensions	Age-adjusted mortality, by cause of death	Weibull prop. hazard		11.5–29.6% ↓ mortality		Mortality reductions strongest for infectious disease related deaths
Cawley et al. (2010)	NHIS 1990–1992, 1994–1996	55+	Household income	BMI, over- and underweight, obesity	2SLS	SS notch (5–7% income loss)		No effects	
Michaud and van Soest (2008)	HRS 1992–2002 (6 waves)	Couples 51–61 at baseline	Liquid and nonliquid wealth	Health index constructed from SAH, conditions, ADL, CES-D, BMI	Dynamic panel GMM	Inheritances when allow contemporaneous effect	No effects		Models w/o unobserved heterogeneity or with too few lags → different estimates
Schmeiser (2009)	NLSY 1979	Low-income 25–43 years	Family income	BMI and obesity	2SLS, 2 SQR, FE	EITC		Females: ↑ BMI and ↑ prob. obese Males: no effects	
Smith (2007)	PSID 1984–1999	Adults	Household income, total and stock market wealth	Onset of new health conditions/disease	Probit		No effects		No effect of either positive or negative wealth change
Moran and Simon (2006)	AHEAD 1993/94	70+	↓ Social Security income	Use of prescription drugs	OLS and IV	SS Notch	–	↓ prescription drug use among low income	

Table 17.5 US evidence of income and wealth effects on adult health and health behavior—cont'd

Authors	Data	Sample	Income/ wealth measure	Measures of health/ health behavior	Estimator	IV for income/ wealth	Effect on health	Effect on health behavior or medical care use	Remarks
Snyder and Evans (2006)	MCOD NHIS 1986–1994	65+	Family income	5-year mortality rates	DID and RDD	SS Notch	↑ mortality		
Adams et al. (2003)	AHEAD 1994–1998	70+	Liquid and nonliquid wealth, household income	Mortality, acute and chronic conditions, SAH, ADL, BMI, smoking	Probit and ordered probit		Non-causality not rejected except for mental health and some chronic conditions	Noncausality rejected only for smoking (males) and BMI (females)	UH not eliminated
Meer et al. (2003)	PSID 1984–1999	Household heads	Net wealth (not pension wealth)	SAH	Two-stage probit	Gifts or inheritances > $10,000	No effect	–	
Deaton and Paxson (2001)	CPS 1976–96	25–85	Equivalent family income	Mortality	OLS and IV	Schooling and cohort dummies	Long term (middle age): ↓ mortality Short-term (young men): ↑ mortality	–	
Ettner (1996)	NSFH 1987, SIPP 1986–1997, NHIS 1988	18–65	Family income	SAH, work and functional limitations, bed days, alcohol consumption, depressive symptoms	OLS and 2SLS	Work experience and state unemp. rate, parental and spousal education	Strong and large ↑ physical and mental health	No effects	Dubious IVs

Note: See Appendix for explanation of dataset, variable, and estimator acronyms.
[a]Effect on health/health behavior is with respect to health measure defined in this column.

further by conditioning on a battery of health conditions that are presumed, in a medical sense, to be precursors of the illness that income (wealth) is hypothesized to affect.

In this study, the null stating that lagged income (wealth) does not predict health is not rejected for most conditions, including acute, sudden-onset conditions and mortality. The authors interpret this result as consistent with the absence of a causal effect of income and wealth on most health outcomes.[39] The hypothesis of no causal effect of wealth is rejected for the incidence of mental health problems, and the results for chronic and degenerative diseases are mixed. The authors argue that, because the treatment of mental and chronic illnesses are often not (fully) covered by Medicare, the ability of the individual to pay for such care may be a causal factor in the determination of these conditions.[40]

Stowasser et al. (2012) revisit the analysis by applying the same Granger causality tests to the original data source extended to a longer observation period, younger cohorts at a given age and younger ages (50+ rather than 70+). The last extension generates some variation in health insurance status that was not present in the older, Medicare-eligible sample included in the original study. With these changes, the null that health is conditionally independent of lagged income and wealth is rejected for a much larger number of conditions, leaving only a minority of conditions for which it is not rejected. This result is problematic for the approach because rejection of the null can arise either due to a true causal effect or a common correlation with omitted unobservables. So, while the original study tends toward the conclusion of no causal effect of income or wealth on health, analysis of more data leaves one in the unfortunate situation of being unable to make any conclusions about the existence of the causal effect.

17.4.2.2 Causal Effects

Panel data methods can be used to deal with time-invariant unobservable determinants of health and income that cloud the conclusions that can be drawn from Granger causality analysis, although this is not straightforward because it uses the nonlinear estimators appropriate for modeling categorical health measures when dynamics and long-term relationships are taken into account. Estimating a dynamic random effects model of SAH with British data, Contoyannis et al. (2004) find that health varies with income averaged over time but not with current income (see Table 17.6 for all studies providing evidence from Europe). This result might be interpreted as indicating that health responds to changes in permanent income but not to transitory income shocks. As such, it is consistent with the above-mentioned argument that sudden income surprises observed over a

[39] Considering that correlation through omitted common determinants is not ruled out, this interpretation of the test outcome is based on the presumption that there is no offsetting bias from unobserved heterogeneity that might confound a true causal effect enough for the net association to be insignificantly different from zero (Heckman, 2003; Stowasser et al., 2012).

[40] Adda et al. (2003) question the plausibility of this interpretation because they find similar results using the same tests applied to Swedish and UK cohorts that are fully covered.

Table 17.6 European evidence of income and wealth effects on adult health and health behavior

Authors	Country and data	Sample	Income/wealth measure[a]	Measures of health/health behavior	Estimator	Effect on health	Effect on health behavior or medical care use	Remarks
Apouey and Clark (2013)	UK BHPS 1997–2005	Adults	Lottery winnings	SAH, mental health (GHQ), physical health problems, smoking, social drinking	FE OLS	Mental health ↑ SAH and physical health—no effects	↑ smoking and drinking	
Van Kippersluis and Galama (2013)	UK BHPS 1997–2008	Adults	Household wealth instrumented by lottery win	Smoking and drinking	IV FE		↑ in moderate, not excessive drinking No effect on smoking	
Adda et al. (2009)	UK FES, GHS, and HSE 1978–2003, HMD 1978–1998	Synthetic cohorts 30–60 years	Equivalent household income (cohort averaged)	Mortality, SAH, chronic illness, blood pressure, cardiovascular/respiratory diseases. Smoking and drinking	GMM	Mortality ↑ Health/morbidity—no effects	Smoking/drinking ↑	Assume no effect of cohort health shocks on income
Gardner and Oswald (2007)	UK BHPS 1996–2003	Adults	Lottery winnings 1000–120,000 GBP	Mental health (GHQ)	OLS	↑ mental health (↓ 1.4/36 GHQ points) 2 years after win		Estimates derived from only 137 wins >1000 GBP
Frijters et al. (2005)	Germany SOEP 1984–2002 (west) and 1990–2002 (east)	18 +	Household income	Health satisfaction (1–10)	FE ordered logit	Significant but small ↑ health; in East Germany is small effect for males only	–	No exogenous income variation for West Germans
Lindahl (2005)	Sweden SLLS 1968, 1974, 1981	Adults	Disposable family income (tax register) instrumented by lottery win	Health index, mortality, overweight	2SLS and probit	10% ↑ in income → ↑ health index 0.04–0.09 SD and ↓ 2–3 ppt prob. death within 10 years	No effect on overweight	IV estimates for mortality by large. No effects for older (60 +)

Continued

Table 17.6 European evidence of income and wealth effects on adult health and health behavior—cont'd

Authors	Country and data	Sample	Income/wealth measure	Measures of health/ health behavior	Estimator	Effect on health	Effect on health behavior or medical care use	Remarks
Contoyannis et al. (2004)	UK BHPS 1991–1999	16+	Household income (current = annual permanent = mean over panel)	SAH	Dynamic RE ordered probit	↑ health larger for permanent than current income than current income. Larger effect for men than women		Cannot separate UH from permanent income in mean income effect
Deaton and Paxson (2004)	England and Wales FES 1971–1998	25–85	Gross income per adult equivalent (cohort average)	Mortality and tobacco expenditure (cohort averages)	OLS	No coherent, stable effects	–	Tenuous identifying assumption of time-invariant age effect on mortality
Jensen and Richter (2004)	Russia RLMS 1995–1996	Pensioner households	Pension income	Mortality, ADL, calorie and protein intake, medication	FE	Pension arrear ↑ 2-year mortality for men by 6%	↓ nutrition, medication, and checkups	
Adda et al. (2003)	Sweden ULF	28–84	Household income	As in Adams et al. (2003)	As in Adams et al. (2003)	Noncausality rejected only rejected for similar conditions to Adams et al	Noncausality rejected from smoking and BMIs	

Note: See Appendix for explanation of dataset, variable, and estimator acronyms.
[a]Effect on health/health behavior is with respect to health measure defined in this column.

short period may not provide variation in economic circumstances relevant to the determination of health. Sustained differences in income that influence long-term behavior seem more relevant to the evolution of health. However, caution is called for because it is not possible to separate the effect of individual income averaged over a panel from that of time-invariant correlated unobservables.

Frijters et al. (2005) exploit the largely exogenous income variation generated by the reunification of Germany in 1990 that resulted in sudden large income gains for nearly all residents of the former East Germany. Reverse causality cannot be eliminated because the East German component of the panel only started in 1990, and so reunification cannot be used as an instrument. Fixed effects models of reported health satisfaction reveal positive effects of income on health in the West, but, surprisingly, in the East where the income variation was much greater, these effects are only observed for males. However, all estimated effects are very small. Taking into account that the estimates are potentially upwardly biased by the failure to eliminate reverse causality, this study suggests that income does not have a substantial causal impact on health (satisfaction) in Germany.

Using data aggregated at the level of birth cohorts, Deaton and Paxson (2001) find strong negative effects of income on all-cause US mortality in the period 1976–1996. The effects appear strongest in middle age and in young men. But these findings are not uncontroversial. It is difficult to rule out reverse causality in cohort models, and the authors' use of education as an instrument for income is easily criticized. Moreover, the same authors do not find any coherent or stable effects of cohort income on cohort mortality in England and Wales (1971–1998) (Deaton and Paxson, 2004). They conclude that the observed correlated cohort income growth and mortality decline in both countries does not necessarily reflect a causal effect of the former on the latter, but it more plausibly arises from technological advances and the emergence of new diseases, such as AIDS, that affect age groups differentially. In this case, the main identifying assumption of the cohort approach—that age effects on mortality are constant through time—is invalid. This rather negative conclusion has not kept others from adopting a similar approach. Adda et al. (2009) study the health effect of permanent income innovations arising from structural changes in the UK economy in the 1980s and 1990s that are assumed to be exogenous. They find that cohort incomes have little effect on a wide range of health outcomes, but they do lead to *increases* in mortality: a 1% increase in income is estimated to lead to 0.7–1 more deaths per 100,000 persons among the prime-aged (30–60) population in any given year. This result is in sharp contrast to Deaton and Paxson's finding of no mortality effect for the UK, and a negative effect of income on mortality for the USA. The authors claim their finding is consistent with substantial evidence that population health is countercyclical (Ruhm, 2000, 2003), although they are identifying the health effect of permanent income shocks, not transitory changes.

The countercyclicality of health is consistent with accumulating evidence from the USA of deterioration in individual health coinciding with the receipt of income. Evans and Moore (2011) find that mortality increases immediately following the arrival of monthly Social Security payments, regular wage payments for military personnel, tax rebates, and dividend payments. The increase in mortality is large and occurs for many causes of death connected to short-term behavior—like heart attacks and traffic accidents—but not for cancer deaths, which suggests that the effects derive from increased risky behavior. For example, the daily mortality of seniors is half a percentage point higher in the week after Social Security pay checks arrive compared to the week before. Mortality in younger populations is even more responsive to income receipt. Dobkin and Puller (2007) find elevated drug-related hospital admissions (23%) and within-hospital mortality (22%) in California in the first few days of the month for recipients on federal DI programs paid on the first of the month.

Health deterioration in response to payment of a given level of income is not necessarily inconsistent with health improvement arising from a permanently higher level of income. Higher income may afford both a smoother consumption profile and a lifestyle that is freer of health-threatening binging on alcohol or drugs. Although the evidence on the health response to the receipt of income rightly makes one wary of the health consequences of increased intermittent cash payments to certain groups, it tells us nothing about how the level of income impacts health.

Identification of the health effect of windfalls arising from prizes, lottery wins, investment returns, or inheritances is attractive because the gains are unanticipated and so are more plausibly exogenous to the evolution of health. Smith (2007) exploits large wealth gains accumulated by US stockholders during the stock market run-ups of the late 1980s and 1990s to estimate effects on the onset of major and minor chronic conditions, while conditioning on baseline health, income, and wealth. He does not deal with unobserved heterogeneity and so uses the language of prediction, not causation. Wealth changes (positive or negative) do not predict health changes.

Using the same PSID data but instrumenting wealth by inheritances, Meer et al. (2003) also find no significant effect on health. The same negative result emerges from three studies that test for a response of health to inheritance-induced changes in wealth using data on older (50+) individuals from the HRS (Carman, 2013; Kim and Ruhm, 2012; Michaud and van Soest, 2008). Allowing for a rich lag structure and unobserved heterogeneity, Michaud and van Soest (2008), as was noted in Section 17.3.7, find a significant effect of health on wealth, but they find *no* evidence of a causal effect of (contemporaneous or lagged) wealth on either SAH or chronic conditions.[41] Carman (2013) finds that health is only correlated with inheritances that are anticipated, the exogeneity of which may be doubted.

In stark contrast, Schwandt (2013) provides evidence of a positive wealth effect on health among relatively well-off retirees observed in the US HRS. In this admirably

[41] Inheritances are only used as an instrument for wealth in the models that test for a contemporaneous effect.

careful and detailed study, the author constructs a claimed exogenous measure of wealth shocks from the rate of change in the S&P 500 stock market index over a 2-year period applied to the household's share of lifetime wealth held in stocks. The use of this measure presumes that it is the proportionate, and not the absolute, change in wealth that potentially impacts health. Thus, a psychophysiological response to a relative change in wealth, perhaps operating through stress, will generate more health variation in relatively wealthy, stock-holding retirees with health insurance than will a change in absolute wealth, given that the absolute change in wealth is not large enough to influence their ability to purchase health-preserving goods, such as medical care. Schwandt (2013) estimates that a shock (positive or negative) corresponding to 5% of lifetime wealth, which is within the range observed in the data, is positively associated with a change of 1–2% of a standard deviation in a variety of health measures, including onset of new health conditions, reported change in self-assessed health, mental health, and, for negative wealth shocks, even survival. Consistent with the hypothesis of a mechanism operating through stress, the study shows a significant impact on hypertension and, to a lesser extent, heart disease, stroke, and psychiatric problems, but no significant effect on conditions that are likely to evolve more gradually, such as diabetes, arthritis, cancer, and lung disease, although this may also be attributable to the lower incidence of some of the latter conditions. A further clue to a possible mechanism is provided by comparing the estimated health effects of a wealth shock with the cross-sectional correlation of health and wealth. For aggregate health conditions, mental health, hypertension, and heart disease, the magnitude of the estimated effect is greater than the respective correlation. For chronic conditions that take longer to develop, the opposite is true. This is consistent with an abrupt change in wealth of the otherwise well-off triggering a health change that is quite different in nature and aetiology from the health differences by wealth observed in that population.

Zooming in on the October 2008 stock market crash, McInerney et al. (2013) present further evidence from the HRS on the impact of large wealth losses on mental health.[42] The crash reduced wealth and increased depressive symptoms, as well as the use of antidepressants. The effects are nontrivial: for instance, a loss of $50,000 in the value of nonhousing wealth is estimated to increase the likelihood of feeling depressed by 1.4% points (8% in relative terms). Although one may expect some recovery from these immediate and substantial declines in mental health as result of adaptation, the evidence from

[42] There is striking and persistent evidence from analyses of aggregated data showing that mortality tends to follow the business cycle, increasing during booms and declining during recessions (Ruhm, 2000, 2003). Although this evidence is not necessarily directly relevant to an explanation of the strong positive cross-sectional correlation between income and health, it does appear inconsistent with income gains causally raising health. The health benefits of recessions have been contested in research covering the post-2007 Great Recession period by McInerney and Mellor (2012), Stevens et al. (2011), Tekin et al. (2013), as well as Ruhm (2013) himself.

Schwandt (2013) would suggest that the mental stress may provoke the onset of risk factors for physical illness.

The main threat to identifying the short-term health effect of a wealth shock arising from share prices is the possibility that health risks differ systematically with the fraction of wealth held in stocks. Schwandt (2013) provides some analysis that suggests this is not driving his results. However, a dynamic model of joint decisions over financial and health investments, which has a good fit with the relationships observed in PSID data, does predict that individuals facing greater health risks will diversify by holding less risky financial investments (Hugonnier et al., 2013). This model also predicts that investments in health rise steeply with wealth. While resting on strong behavioral assumptions, it provides insight into the joint evolution of health and wealth and could potentially be useful for pinpointing strategies for convincingly identifying the wealth effect on health.

A few European studies find evidence of positive health effects resulting from lottery wins. Using a Swedish panel and instrumenting a measure of permanent income (average income over 15 years) with average lottery winnings, Lindahl (2005) estimates that an income increase of 10% generates a fall in morbidity and a rather spectacular 2–3% point decrease in the probability of dying within 5–10 years. One may be sceptical of the credibility of such a large effect, which exceeds even the raw correlation between income and mortality. Using British data, Gardner and Oswald (2007) find that 2 years after a win of between £1000 and £120,000, the GHQ index of mental health increased by 1.4 points, on a scale of 36 points. The effect is only significant for males and, surprisingly, for higher-income individuals. Using a few more waves of the same data, Apouey and Clark (2013) find that lottery winnings have no significant effect on SAH, but a large positive effect on mental health.

Although the exogeneity of windfalls is certainly valuable, one may question the relevance of the resulting evidence to an understanding of the large differences in morbidity and mortality between the rich and the poor that are likely to arise from sustained differences in health behavior, and perhaps access to medical care, over many years. Economic shocks observed in data with a limited longitudinal span are potentially useful in identifying short-term health responses, but they can tell us little or nothing about the mechanisms responsible for the gradient in health conditions that emerge over the life cycle.

Pension policies have provided a final source of income variation from which researchers have attempted to identify effects on health. Jensen and Richter (2004) study the effect of losses in pension income in Russia during a major crisis period (1995–1996) (Table 17.6). Delayed pension payments had a dramatic impact on living standards, with income declining by up to 24% and poverty rates tripling to over 50%. For males, the loss of pension income increased the likelihood of death within 2 years by 5.8% points, and raised functional impairment (ADL) and the probability of experiencing chest pain. These effects are likely to have materialized from substantial and significant reductions in both calorie and protein intake, as well as reduced use of medication for chronic conditions and preventive checkups. There were no effects on women's health or mortality.

Back in the USA, Snyder and Evans (2006) report evidence suggesting that reduced pension income *raises* health (Table 17.5). They exploit a notch in Social Security payments that resulted in those born after January 1, 1917, receiving sharply lower retirement incomes than contemporaries with identical earnings histories born slightly earlier. There was little time to adjust to the income loss because the legislative changes happened late in their working lives. In any case, most of those affected did not realize the impact of the changes until after retirement. The authors find that the decrease in pensions reduced mortality and rationalize this surprising result by a claimed positive health effect of increased postretirement (part-time) work effort in response to the income loss.

The contradictory evidence from the USA and Russia is most plausibly attributable to differences in how the level of income around pensions fluctuated. There is evidence of large mortality reductions (particularly for the poor) resulting from increases in pensions paid to US Union Army veterans at the beginning of the twentieth century (Salm, 2011), when both incomes and health were obviously much lower than in the period studied by Snyder and Evans.

17.4.3 Income and Wealth Effects on Health Behavior

The evidence reviewed in the previous section does not support a strong, or even any, causal effect of income, or wealth, on health. But this may simply reflect the difficulty of observing, even in moderately long panels, the health consequences of changes in health behavior and utilization of medical care that may only materialize in the long term. In this and the following section, we assume that there are health effects of smoking, drinking, forgoing effective health care, and so on, and we examine whether there is evidence that economic circumstances impact on these health determinants.

Particularly in the USA, more affluent individuals are generally less likely to smoke, drink heavily, be overweight, or use illegal drugs, and they are more likely to exercise and engage in preventive care (Cutler and Lleras-Muney, 2010; Cutler et al., 2011a,b). But simple correlations obviously tell us nothing about the presence or direction of causality. Some of the evidence reviewed by Cawley and Ruhm (2011) shows that income and/or wealth increases consumption of tobacco and alcohol. This holds for the response to income shocks in the UK captured by cohort income (Adda et al., 2009) and lottery winnings (Apouey and Clark, 2013) (see Table 17.6). In the USA, Kim and Ruhm (2012) find that wealth gains from inheritances only raise moderate drinking and have no effect on smoking (Table 17.5).

The evidence for income and wealth effects on obesity is mixed but certainly does not support a strong causal effect in either direction. Kim and Ruhm (2012) find some indication of wealth gains reducing the likelihood of being overweight, which is consistent with Swedish evidence based on lottery winnings (Lindahl, 2005). Cawley et al. (2010) use US National Health Interview survey data and the Social Security notch as an IV for income and find no impact of income on weight or obesity. Exploiting variation across US states in the generosity of the Earned Income Tax Credit (EITC), Schmeiser (2009) finds no effect of income on weight for men and a positive effect for women: an additional \$1000 per year is associated with a gain of no more than 1.80 pounds (0.82 kg).

Galama and van Kippersluis (2010) extend Grossman's (1972a,b) health-capital model with the aim of understanding how health behavior may differ by wealth. They distinguish between healthy consumption, which reduces the rate of depreciation of health (e.g., good housing, vitamins, muesli) and unhealthy consumption, which increases health depreciation (e.g., cigarettes, excessive alcohol, etc.). Wealth has a positive effect on healthy consumption both because of a pure wealth effect and because higher wealth raises health investment, which is assumed to exhibit diminishing returns, so that depreciated health is more expensive to replace through medical care.[43] The effect on unhealthy consumption is ambiguous because the wealth and price effects go in opposite directions. The wealthy are less inclined to run down their health because of the higher marginal cost of replacement. The model predicts that, under arguably plausible assumptions, the wealthy will be more likely to engage in moderately unhealthy consumption (wealth effect dominates) and less likely to partake in severely unhealthy consumption (price effect dominates).

Van Kippersluis and Galama (2013) test these predictions with wealth gains instrumented by lottery winnings in British (BHPS) data (Table 17.6), as in Apouey and Clark (2013), and by inheritances in US (HRS) data (Table 17.5), as in Kim and Ruhm (2012). Unlike the earlier studies, they use fixed effects models to deal with unobserved heterogeneity and find robust evidence that wealth increases the probability of drinking alcohol, but it has no effect on the number of drinks and heavy drinking. This is consistent with the direct wealth effect dominating for behavior that is moderately unhealthy, although indulgence in a glass of good claret over dinner may actually be beneficial to one's health. Their results for smoking are inconclusive: a lottery win in the UK does not increase smoking, which is inconsistent with Apouey and Clark (2013), but inheritance receipt in the USA does immediately increase both the prevalence and intensity of smoking, which is inconsistent with Kim and Ruhm (2012), who look at longer term effects, and also with the prediction of the theory.

Income opens consumption opportunities. For unhealthy consumption to explain income-related health inequality, tobacco, alcohol, fatty foods, and so on would have

[43] The assumption that health investment technology exhibits diminishing returns is a departure from Grossman's model that assumes constant returns. With diminishing returns, because greater wealth raises the demand for health and the level of investment, the marginal cost of producing a unit of health by investing in medical care is greater at higher levels of wealth. This higher marginal cost gives the wealthy an added incentive to look after their health.

to be grossly inferior goods. There is no evidence of this. This is not to say that health behavior is not an important contributor to the social, as opposed to economic, gradient in health. On the contrary, health behavior can account for a large proportion of the differences in health across education groups (Cutler and Lleras-Muney, 2010; Cutler et al., 2011a,b). But it is likely to be the preferences and knowledge of higher education groups, and not their wealth, that lead them to adopt healthier lifestyles.

17.4.4 Income Effects on Medical Care

As pointed out earlier, the potential for medical care to contribute to health differences by income is constrained in many high-income countries by the dominance of public health insurance. The income gradient in utilization of medical care should be stronger in countries, such as the USA, that give the market greater more influence on the financing of the health system. But public health care seldom completely crowds out private care, and even within the European social health insurance and national health service systems, specialist care is often distributed in favor of the better-off (Van Doorslaer et al., 2000, 2004, 2006). But while the income elasticity of demand for medical care has been the subject of numerous studies, the literature provides surprisingly little evidence of a causal effect of income on utilization.[44]

The Kim and Ruhm (2012) study using the US HRS finds that wealth gains from inheritances raise utilization of many types of medical services and out-of-pocket spending. Using a sample of the old (70+) US population whose drug expenses were not covered by Medicare at the time, Moran and Simon (2006) find a large and statistically significant effect of income instrumented by the Social Security notch on prescription drug utilization, though only for households that have low education and do not have high income (<75th percentile) (Table 17.5). Their estimates of income elasticity are all above 1. Goda et al. (2011) extend the analysis to estimating the impact of income on utilization of long-term care, which is also not fully covered by Medicare, and they find that a positive permanent income shock lowers nursing home use but increases the utilization of paid home care services. It is important to bear in mind that the estimated positive income effects on both drug and long-term care utilization pertain only to the elderly who had been low-wage workers, because the Social Security notch had only a weak impact on the pensions of older cohorts who had been higher earners.

These three studies of the older US population confirm what one would expect. Medical care is a normal good. Where universal public health insurance coverage is absent, individuals who can afford more and better health care will purchase it.

[44] Virtually every textbook in health economics devotes a chapter to the demand for medical care, including estimates of income elasticity (e.g., Sloan and Hsieh, 2012). We do not cover the literature on the effect of income on the demand for health insurance, which would take us some distance from the income–health nexus. It is well-known that uptake of insurance is very much income-related.

17.4.5 Income Effects on Child Health

In Section 17.3.4, we conclude that early-life health conditions have an economically significant effect on economic well-being in adulthood. Currie (2009) proposes that child health is a potentially important contributor to the intergenerational transmission of education and economic status. The idea is that less educated, poor parents are more likely to give birth to and rear less healthy children. Childhood ill-health interferes with human-capital acquisition and directly constrains health capital in adulthood, which further reduces earnings potential. A cycle of poverty is propelled by childhood ill-health. Poverty begets childhood illness, which generates poverty later in life. If true, this would give health a role not only in the creation of inequality, through health shocks that increase income dispersion, but in its perpetuation across generations. Whether parental income does constrain child health is therefore an important question to be addressed not only from a health perspective but also from that of economic inequality.

We focus here on the evidence that parental economic circumstances constrain child health and skip consideration of the mechanisms through which an effect may arise. Almond and Currie (2012) use Cunha and Heckman's (2007) model of investment in cognitive and noncognitive skills of children to provide a framework for thinking about the evolution of children's human capital, including health. Further development of this model to formally incorporate health may provide insight into the impact of parental income on child health.

From a review of the evidence, Currie (2009) concludes that, while there is little doubt that children from less privileged backgrounds are less healthy, there is insufficient evidence to conclude that, in a high-income country context, this arises from a causal effect. Identification of a causal effect of parental income on child health should, nonetheless, be easier than the identification of the (own) income effect on adult health. The reason is that reverse causality is less of an issue because children generally do not earn income, at least in high-income countries. This has been one of the main motivations for researchers to examine the impact of parental income on child health. Of course, reverse causality is not entirely eliminated, because the illness of a child may interfere with his or her parents' work activity, and correlated unobservables remain a substantial problem.

17.4.5.1 Evidence from High-Income Countries

The correlation between family income and children's general health strengthens as children grow older in the USA (Case et al., 2002) and Canada (Currie and Stabile, 2003), suggesting that the disadvantages associated with parental income accumulate as children age. The steepening of the gradient with age can be due to poorer children being hit by more health shocks and/or having more difficulty recovering from illness, given constrained access to medical care. In the USA, the strengthening of the gradient is due to a combination of these effects (Case et al., 2002; Condliffe and Link, 2008), whereas in Canada, consistent with its universal health care system, it is only due to poor children becoming sick more frequently (Currie and Stabile, 2003).[45]

These findings are not generally confirmed for other countries. Khanam et al. (2009) find that there is a gradient in Australia that strengthens with age when similar covariates to those used by Case et al. (2002) are included. However, the gradient disappears when they include a richer set of controls, in particular maternal health, suggesting that there may be no causal effect. Reinhold and Jürges (2012) find that the parental income gradient in child health in Germany is as strong as it is in the USA, but it does not steepen as children grow older, which could be attributed to the constraining effect of universal health care.

The UK evidence is mixed, with Currie et al. (2007) and Case et al. (2008) arriving at different conclusions from analyses of the same survey. Currie et al. (2007) find a significant family income gradient in child general health that increases between ages 0–3 and 4–8 and decreases afterward. Case et al. (2008) add 3 years of data and find that the gradient keeps increasing until age 12. Analyses of a rich data set from one region of England reveal a gradient that does not increase between birth and age 7 and almost disappears with an expanded set of controls, including parental behaviors and health (Burgess et al., 2004; Propper et al., 2007) Using a nationally representative sample, Apouey and Geoffard (2013) find a gradient that persists up to the age of 17 but no evidence that utilization of health care, housing conditions, nutrition, or clothing are important mechanisms for generating it.

In North America and Europe, children from poorer households are less healthy. Whether this arises from an effect of parental income or some other characteristic of the family associated with both income and child health cannot be established from the studies cited above. Using reform-induced variation in the US Earned Income Tax Credit (EITC), Hoynes et al. (2012) estimate that increased maternal income reduces the incidence of low birth weight and increases mean birth weight. For single low-educated mothers, an increase of $1000 in the EITC generates a 6.7–10.8% reduction in the incidence of low birth weight (see Table 17.7 for this and other studies providing causal evidence in this and next subsection). The effect appears to be mediated through slightly greater use of prenatal care and much more substantial reductions in smoking and drinking during pregnancy.[46] These estimates suggest sizeable gains in infant health from income increases among low-income populations.[47] A much more

[45] Allin and Stabile (2012) find no evidence that health care utilization is an important factor for generating the gradient in Canada.

[46] A $1000 credit received by a low-educated single mother is estimated to increase the propensity to use prenatal care by 0.65% points (from a baseline of 96%) and to reduce the likelihood of smoking by 1.2% points (baseline of 30%) and of consuming alcohol by 1.1% points (baseline of 3.3%). It is not clear why increased income reduces smoking and drinking, although one might suppose that it has to do with reduced financial stress.

[47] There is also evidence of the child health impact of targeted programs such as food stamps (Almond et al., 2011) or food and nutrition vouchers (Hoynes et al., 2011). These are not considered here because of their conditional nature, although Hoynes and Schanzenbach (2009) claim that recipients of food stamps behave as if the benefits were paid in cash.

Table 17.7 Evidence of income effects on child health and related health behavior

Authors	Country and data	Sample	Income measure	Measures of health/health behavior	Estimator	Effect on health	Effect on health behavior and medical care	Remarks
Mocan et al. (2013)	US Natal Detail Files and CPS Annual Demographic Files 1989–2004	Singleton births to unmarried mothers >19 years	Weekly earnings instrumented by measure of skilled-biased technology shocks	Birth weight, gestation age, prenatal care, smoking and drinking during pregnancy	Two sample IV	Small ↑ birth weight and gestational age babies of low education (≤high school) mothers unlikely to be on Medicaid	↑ prenatal care for low educ. mothers not on Medicaid No effect on smoking or drinking	Effects are very small. Doubling of income ↑ birth weight by 100 g and ↑ gestational age by 0.7 weeks No effects for high education mothers or those likely to be on Medicaid
Hoynes et al. (2012)	US Vital Statistics Natality Data 1983–1999	Infants of low-income mothers	Change in maternal income due to EITC reform	LBW, smoking, drinking, prenatal care use	DID	\$1000 → 6.7–10.8% ↓ LBW	\$1000 → prenatal care ↑ 0.65 ppt smoking ↓ 1.2 ppt drinking ↓ 1.1 ppt	
Amarante et al. (2011)	Uruguay mortality, natality, Social Security registers 2003–2007	All newborns	Unconditional cash transfer = 50–100% prepayment income to low-income women	LBW, maternal smoking, nutrition	DID, FE, and RDD	15% ↓ LBW (1.5 ppt of baseline 10 ppt)	↑ maternal nutrition, ↓ smoking during pregnancy, ↓ % children born to unmarried parents	Transfer paid women in bottom income decile
Fernald and Hidrobo (2011)	Ecuador, 2003–2006	Children 12–35 months	Unconditional cash transfer = \$15 per month (6–10% mean household exp.) to low-income women	HAZ score, hemoglobin concentration, vitamin A and iron supplements	Randomized experiment	No effects	↑ vitamin A and iron supplements in rural areas	Transfers paid to women in poorest 40%
Agüero et al. (2009)	South Africa KIDS 1993, 1998, 2004	Children	Unconditional cash transfer paid to women	HAZ scores	GPSM	Large cash transfer early in life ↑ HAZ		

Table 17.7 Evidence of income effects on child health and related health behavior—cont'd

Authors	Country and data	Sample	Income measure	Measures of health/health behavior	Estimator	Effect on health	Effect on health behavior and medical care	Remarks
Paxson and Schady (2010)	Rural Ecuador 2003–2006	Children 36–83 months	as Fernald and Hidrobo (2011)	Hemoglobin level, HAZ, fine motor control, deworming treatment	Randomized experiment	Poorest quartile: ↑ hemoglobin Other quartiles: no effects	↑ deworming treatments	as Fernald and Hidrobo (2011)
Case (2004)	South Africa Langeberg Survey 1999	Children	Pension receipt	height	OLS	↑ height of black and colored kids		
Duflo (2000, 2003)	South Africa SALDRU 1993	Children 6–60 months	Woman's receipt of pension	HAZ and WHZ scores	2SLS (IV = pension eligibility)	↑ HAZ and WHZ of girls by 1.2 SD after 2 years		No effect on boys. No effect of pension paid to male on either girls or boys.

Note: See Appendix for explanation of dataset, variable, and estimator acronyms.

modest effect is estimated from data on 14 million US births between 1989 and 2004 that uses a census division year-specific index of skill-biased technological change to instrument mothers' earnings (Mocan et al., 2013). For low-educated (i.e., no more than high school diploma) unmarried mothers who are unlikely to be on Medicaid (public health insurance for low-income households), increased earnings raise utilization of prenatal care, as well as birth weight and gestational age. The fact that there are no significant effects on births to high-educated mothers and to all mothers who are likely to be covered by Medicaid suggests that low income constrains access to maternity care for those lacking insurance coverage. However, the effects are very small. A doubling of earnings would raise birth weight by only 100 grams and gestational age by only two-thirds of a week.

17.4.5.2 Evidence from Low- and Middle-Income Countries

One would expect health in general, and child health in particular, to be more contingent on income in low-income settings where the nutritional needs to sustain health are often not met and universal health insurance coverage is absent, with most medical care paid for out-of-pocket. Indeed, the economic gradient in health is particularly steep in low- and middle-income countries and is evident in critical indicators, such as infant mortality (Commission on the Social Determinants of Health, 2008). The evidence that the gradient in child health does derive, at least in part, from the causal impact of economic circumstances on health is much more clear-cut than that from high-income countries (see Table 17.7 for studies cited in this section).[48]

Duflo (2000, 2003) examines whether the extension of pensions to black South Africans in the early 1990s had an impact on the nutritional status of children. An effect may have been anticipated because more than a quarter of black children under five lived with a pension recipient by the end of the period studied. The analysis reveals that pensions paid to women have substantial positive effects on the weight and height of girls but no significant effects on the nutritional status of boys, and pensions paid to males have no effect.[49] The effects are very large. Payment of a pension to a woman

[48] A number of studies have exploited macroeconomic shocks to identify the impact of income on health and (infant) mortality in nations of the developing world, including Mexico (Cutler et al., 2002), Peru (Paxson and Schady, 2005), India (Bhalotra, 2010), Colombia (Miller and Urdinola, 2010), and 59 other countries (Baird et al., 2011). Although these studies are able to identify health effects at the individual level, they are unable to trace the income consequences of the macro shocks at this level. They tend to find substantial negative effects of aggregate income on mortality, but tell us little about the extent to which variation in income across individuals generates inequality in health.

[49] The effect on weight-for-height z-score, which should respond immediately to improved nutrition, is identified by comparing children living in households with elderly relatives eligible for pensions (>59 for females and >64 for males) with others with older relatives that did not quite reach the age of pension qualification. The effect on height-for-age z-score, which reflects longer term nutritional intake, is identified by comparing the height deficits of younger and older children living in households with an elderly person eligible for a pension relative to those in other households. A smaller deficit among younger children is consistent with a positive impact of income on height because the younger children lived in households benefiting from the pension extension for a larger proportion of their lives.

is estimated to raise both weight and height of girls by 1.2 standard deviations over a 2-year period (Duflo, 2003). The income gain was also large with pension benefits being around twice the median per capita income in the rural areas at the time. These results suggest that income can have very large positive effects on child health in low-income settings, but whether this effect materializes crucially depends on who receives the income.[50] Consistent with this, an unconditional cash grant paid to child caregivers (mostly women) in South Africa has been demonstrated to significantly boost child height (Agüero et al., 2009). On the basis of the observed relationship between adult height and earnings, the projected discounted return to the grant is estimated to be as much as 50%.

The evidence is mixed for income effects on child health in poor populations, as obtained from unconditional cash transfer programs in Latin America, however.[51] Ecuador's *Bono de Desarrollo Humano* (BDH) pays $15 per month—equivalent to 6–10% of average household expenditure in the target group—to mothers of children below the age of 17 in the poorest two-fifths of the population, but this additional income has been found to have no significant impact on the health (height and hemoglobin concentrations) of children aged 1–3 years (Fernald and Hidrobo, 2011), and among older children aged 3–7 years, there are only modest effects (on hemoglobin and deworming treatments) for the poorest (Paxson and Schady, 2010).[52] Uruguay's PANES program targets poorer households, restricting payment of the generous monthly cash transfer—equivalent to 50% of average preprogram income for recipient households and up to 100% of income for households with a recent birth—to households in the bottom decile. It is estimated to reduce the incidence of low birth weight by 1.5% points relative to a baseline of 10% (Amarante et al., 2011). This effect appears to materialize through improved maternal nutrition, reduced smoking during pregnancy, a large reduction in the proportion of children born to unmarried parents, and a modest reduction in maternal labor supply. The larger health impact relative to that of the general cash transfer in Ecuador is plausibly explained by the greater magnitude of the payment and its direction to (relatively) poorer households.

[50] The health gains from pension income in South Africa are not confined to children. Case (2004) finds that the extra income brought by the presence of a pension in a household (equal to 2.5 times the median income in the sample analyzed) improves the health status of all adults in households in which income is pooled.

[51] Conditional transfer schemes are less interesting for our purpose because payments are made conditional upon behavior (e.g., school attendance, medical care receipt, attendance of preventative health services, health, and nutrition education) that are intended to have a direct impact on health. While many of these programs have proven highly effective, Gertler (2004), Rivera et al. (2004), and Fernald et al. (2008) have shown it is difficult to separate the pure income effect from the incentive effect.

[52] In rural areas, vitamin A and iron supplementation did increase, and language development improved among children aged 1–3 years (Fernald and Hidrobo, 2011).

17.4.6 Conclusion

In this section, we set out to determine the extent to which differences in economic circumstances contribute to health inequality across individuals. Do the poor experience worse health because they are poor? Answering this question is difficult because worse health would be expected to be associated with lower income even without being caused by it. This has driven researchers to search for phenomena that generate variation in income or wealth without being caused by or associated with health. In high-income countries, this research enterprise has tended to produce evidence indicative of no impact of income on health in adulthood, or effects that are small in comparison to the observed income-health gradient, suggesting that the association does not derive from financial resources impinging on health. An exception is US evidence of deterioration in health, particularly as reflected in indicators related to stress and mental status, in response to stock market losses. To an extent, the general finding of little or no effect is plausible. Variation in health arises from differences in the health stock with which we are endowed (genetics), the extent to which we look after this endowment (lifestyle and living conditions), the opportunities to repair it when it gets damaged (medical care), and luck. Financial resources cannot influence the first and last determinants. Most high-income countries offer universal health care coverage irrespective of ability to pay, which greatly weakens the economic impact on the third determinant. That leaves lifestyle and living conditions. Most research concentrates on the former, and, within this, on what we do that is bad for our health, rather than what we do that is good for it. It would be perverse if greater ability to afford indulgencies in unhealthy behavior, such as smoking and drinking, explained why the better-off are in better health. In fact, the rich tend to lead less unhealthy lives, but that is not because they are economically privileged. More likely, it has to do with their education advantage.

The ability of money to buy health in the developed world is limited. Mental health appears to respond to economic circumstances, with losses producing larger deteriorations in mental well-being than gains generate improvements. But there is little evidence that physical health problems are provoked by worsened personal finances. However, we suspect that there is much that current research is missing.

Identification of the effect of one stock variable (financial wealth) on another (health capital) is far from easy. There is a risk that identification strategies that focus on very local effects of windfall gains from lottery wins, inheritances, or tax/benefit reforms throw away effects that accumulate over the life cycle together with the bathwater containing the common unobservables. The determinant that is more permanent—living conditions—tends to get overlooked in research conducted by economists. This includes housing and features of the built and social environment that vary with the economic status of neighborhoods: pollution, leisure facilities, open spaces, food quality, and crime. Money can afford improved housing quality and relocation but it takes a very large economic shock to achieve this. Chronic poverty can entail damp walls, confined spaces, disruptive neighbors, polluted air, and a threat of violence that gradually, or perhaps suddenly, take a toll on health.

The health experience of the chronically poor has no influence on the estimated impact of wealth on health in some research, such as that identifying variation in inheritances or stock prices. Other evidence, such as that based on lottery wins, does potentially capture exposure among the poor, but the sudden and often moderate gains in cash involved may not be sufficient to substantially change living conditions, and even if they are, the observation period is unlikely to be sufficiently long to detect impacts on chronic health problems that may only slowly respond to material circumstances. We are hesitant to conclude that lack of evidence of an impact of wealth on adult physical health in much of the developed world means that there is no effect.

One can be more confident that the worse health of poorer children, which is unfortunately still observed in many high-income countries, is not simply a reflection of health constraining earnings. It may also arise from the fact that poorer parents are also less educated, and this lower level of education impacts child health. As would be anticipated, the strongest evidence that economic conditions determine inequalities in (child) health comes from the developing world. But even here more income does not necessarily bring better health. Money may be able to buy health when nutritional status is low and many cannot afford medical care, but the money must be given to those that value health highly. There is some evidence that women prioritize child health more than men.

17.5. ECONOMIC INEQUALITY AS A DETERMINANT OF HEALTH

17.5.1 Overview

More than 20 years ago, Wilkinson (1990, 1992) introduced the hypothesis that income inequality is harmful to health. He showed that countries with higher-income inequality have lower life expectancy, and others soon confirmed a negative association with other measures of population health (Steckel, 1995; Waldmann, 1992). According to one variant of the hypothesis, this cross-country association reflects a causal effect of income inequality on individual health via psychosocial mechanisms: striving to keep up with the Joneses in societies with higher levels of income inequality raises levels of stress. Income redistribution can potentially raise average health not only because of any greater responsiveness of health to income at lower levels of income, if indeed there is a causal effect of income on health, but also because narrower disparities in income are good for everyone's health, including that of the rich.

The validity of this hypothesis has been heavily debated (Gravelle, 1998; Smith, 1999; Wilkinson and Pickett, 2006), and associated claims that economic inequality is responsible for a host of societal ills beyond poor health, including violence, teenage pregnancy, obesity, mistrust, and high incarceration rates (Wilkinson and Pickett, 2010), has attracted much attention and been subject to a good deal of criticism (Saunders and Evans, 2010; Snowdon, 2010). In the present context, establishing whether there is a health cost of economic inequality is relevant to the evaluation of the strength of an instrumental argument for reducing inequality.

We focus here on the mechanisms through which income inequality can potentially impact health, and we pay close attention to whether empirical analyses are capable of testing the hypotheses. We restrict our attention to the impact of inequality on morbidity and mortality, and we neglect studies of homicides, for which there is general agreement on the importance of income inequality (Deaton, 2003; Lynch et al., 2004b). We also steer clear of the happiness literature that has paid a great deal of attention to income inequality (Alesina et al., 2004; Clark et al., 2008).

In the next subsection, we demonstrate the stylized fact that population health is negatively associated with income inequality, and we outline the mechanisms through which income inequality might threaten the health of all individuals. We then consider alternative theories that can explain the negative association between population health and income inequality at the aggregate level without inequality being a threat to the health of all individuals in a society. We then turn to the evidence.

17.5.2 Basic Hypothesis: Inequality Threatens Everyone's Health

Population health rises with per capita income but at a decreasing rate (Preston, 1975). Among high-income countries, where this so-called *Preston curve* flattens out, population health has been found to be negatively correlated with income inequality (Wilkinson, 1992, 1996; Wilkinson and Pickett, 2010). There is only weak evidence of this in Figure 17.5, which is based on the same data on life expectancy, as well as measures of inequality employed in a popular publication that advances the inequality hypothesis (Wilkinson and Pickett, 2010). In these data, the relationship appears to be driven by the low inequality and high life expectancy of Japan and Sweden, and the high inequality and low life expectancy of the USA and Portugal. Among the bulk of countries with life expectancies of 78–80 years, there appears to be no relationship with income inequality.[53]

A negative correlation between income inequality and average health, presuming it exists, has been attributed to the falling potency of further, material gains in generating health once the average standard of living reaches a threshold beyond which income differences become more relevant to the determination of health (Wilkinson and Pickett, 2010).[54] Two causal mechanisms through which income inequality may threaten the

[53] Some have criticized the criteria used by Wilkinson and Pickett (2010) to select the countries included in the figure, and to exclude others (Saunders and Evans, 2010; Snowdon, 2010). The authors defended their country selections in the second edition of their book, however (Wilkinson and Pickett, 2010).

[54] The literature tends to presume that there is a positive effect of the income level on health.

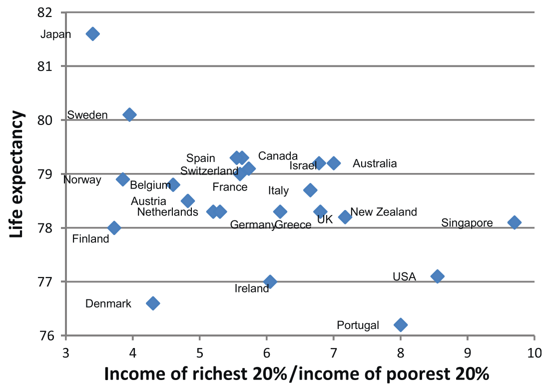


Figure 17.5 Life expectancy and income inequality in high-income countries. *Notes: Data are from Wilkinson and Pickett (2010) who, in turn, took them from UN Human Development Reports. Life expectancy is at birth averaged over males and females in 2004. Income inequality is measured by the ratio of income received by the richest 20% of households to the income of the poorest 20% averaged over the years 2003–2006.*

health of all individuals—rich and poor—have been proposed. The first stresses the importance of the public provision of health-determining goods, while the second focuses on social capital. Some stress a third psychosocial mechanism mentioned above (Wilkinson, 1992). Given that this theory does not propose that income inequality affects health throughout the distribution of income, but rather that the health of less well-off individuals suffers because of their relative deprivation, we cover this assertion in the next section.

The public provision of goods would create a pathway from income inequality to individual health if inequality impacts on the provision of goods that determine health, such as curative and preventive health care, education, and sanitation. Income inequality might lead to more heterogeneous preferences, which will reduce the average value (and thus the provision) of publicly provided goods (Alesina et al., 1999; Deaton, 2003; Thorbecke and Charumilind, 2002). But income inequality could also lead to increased public provision because a more skewed income distribution will reduce the income of the median voter relative to the mean and increase the redistributive effect of public provisions financed by nonregressive taxation (Meltzer and Richard, 1981). Following Sen (1999), Deaton (2003, forthcoming) argues that the focus should be on the health consequences of political, rather than economic, inequalities. He notes that in nineteenth-century Britain, and in the USA and India in the twentieth century, substantial improvements in public health were realized after the extension of political rights.

Social capital—cohesion and trust among citizens (Putnam et al., 1993)—is argued to be a consequence of economic inequality and a determinant of health via social and psychosocial support, informal insurance mechanisms, and information diffusion (d'Hombres et al., 2010; Kawachi and Kennedy, 1997; Kennedy et al., 1998; Ronconi et al., 2012). This hypothesis has received a fair amount of attention in the literature, but whether this attention is justified depends on the assumption that income inequality reduces social cohesion, as opposed to lower social cohesion raising income inequality, which is an equally plausible explanation.

17.5.3 Alternative Hypotheses: Health Responds to Absolute or Relative Income

A negative association between population health and income inequality could arise from the dependence of health on absolute or relative income, without this necessarily implying that inequality threatens everyone's health (Wagstaff and van Doorslaer, 2000). The absolute income hypothesis states that diminishing health returns to income at the individual health level explain the negative association between average population health and income inequality at the aggregate level (Gravelle, 1998; Gravelle et al., 2002; Rodgers, 1979; Wildman et al., 2003). If the health-income relationship is concave then an increase in the spread of the income distribution will bring down mean health because the health loss to those becoming poorer is larger than the health gain to those becoming richer. Income redistribution could raise average population health but this would occur without the health of any individual, given his or her income, being directly affected by the level of economic inequality in society. The literature reviewed in Section 17.4 provides only limited evidence of a causal impact of income on health in high-income countries, but some evidence indicates that the effect is stronger among the poor, and there does appear to be an income effect in low-income countries. If there are health returns to income, they would appear to be diminishing. But even if there is no (diminishing) causal effect of income on health, a negative statistical association between average health and income inequality will be observed when there is a concave statistical relationship between health and income across individuals.

According to the relative income hypothesis, one's health depends on how one's income fares relative to others. When there are diminishing health returns to the difference between individual income and some aggregate, such as the mean, there will be a negative association between average health and income inequality. These hypotheses are distinguished—unfortunately, often not explicitly—from the above-mentioned income inequality hypothesis by the assertion that income inequality only matters to the extent that it increases the number of individuals who have an income deficit relative to some reference level. It is only the health of these individuals that is claimed to be damaged by inequality. In contrast, the income inequality hypothesis postulates that income inequality is a common factor impacting on the health of everyone.

Health is presumed to depend on relative income because of psychosocial effects. It is not so much possession of more material goods that matters for health, but rather the stress, depression, anxiety, shame, and distrust brought on by judging one's standard of living to fall short of that enjoyed by others. These emotional responses are claimed to trigger health-damaging psychoneuroendocrine reactions, such as increased cortisol production (Wilkinson, 1992). Some have even hypothesized that this psychosocial–biological effect may be hardwired into humans through our evolutionary experience (Wilkinson, 2001). Hunter-gatherer societies were extremely egalitarian, and humans might not yet be well adapted to the social inequalities that have arisen in settlement societies. Consistent with the psychosocial–biological mechanism, experiments have found that manipulating baboons, which also have stable, although obviously much simpler, hierarchical societies, into lower social positions induces stress (Sapolsky, 2005). It has been hypothesized that low relative economic status may impact negatively on health through epigenetic responses, as well (Wilkinson and Pickett, 2010).[55] Wilkinson and Pickett cite evidence of the maternal nursing behavior of rats affecting the offspring's epigenome at a (glucocortoid) receptor known to regulate stress responses (Weaver et al., 2004). They propose that, if this also occurs among humans, then increased stress and cortisol levels experienced throughout life could be due to early-life epigenetic processes. These authors do not elaborate on why exposure to such processes may be related to low (parental) socioeconomic position, although there is evidence of associations between epigenetic differences and SES, which is proxied by income position, occupation, education, and housing tenure among adults (McGuinness et al., 2012) and children (Borghol et al., 2012). Understanding of the epidemiological implications of epigenetic processes is still limited (Relton and Davey Smith, 2012), and it would certainly be premature to presume that they contribute to health differences across socioeconomic environments.

Why psychosocial responses should be confined to inequality in the income dimension has been questioned (Deaton, 2003, 2013). It seems quite conceivable that they might be triggered, perhaps more strongly so, by comparisons of occupation, education, housing, and so on. While describing the mechanism in the previous paragraphs, we (and the authors cited) have often resorted to terms such as social hierarchy. The most influential research on health and social position—the Whitehall studies (Marmot et al., 1978, 1991)—has used occupational grade as the discriminating indicator. The evidence obtained from animal studies cannot, of course, inform us about effects in the income dimension. In reviewing the evidence below, we focus on the health effects of relative income, but this is because of the context of this chapter and not because we believe other dimensions of SES to be of lesser importance to health.

Less frequently cited than the psychosocial mechanism is the idea that pecuniary externalities, arising from the pricing of health-enhancing goods, reduce average health in regions that are more unequal (Miller and Paxson, 2006). Take the case of healthy

[55] Although still very much in its infancy, epigenetics is the field of medical science that studies (possibly heritable) random or environment-induced changes in gene expression that are not driven by changes in the underlying DNA sequence (Ebrahim, 2012).

foods. When the quality and availability of healthy food is comparable across regions, but more expensive in rich areas, then poor individuals in these areas will have worse health than their equivalents in poorer areas. Health depends negatively on the individual's income deficit from the regional average because of the price effect on the cost of maintaining health. But there could be an offsetting effect through collectively and locally financed health-enhancing goods, which may include some medical care. The larger tax base of wealthier neighborhoods will increase the supply and quality of health care, which will raise the health of a poor person in the rich region compared with his equivalent in a poor region.

There are three variants of the relative hypothesis—relative income, relative deprivation, and relative position—distinguished (again, often not explicitly) by the functional form linking health to income differences (Wagstaff and van Doorslaer, 2000). The *relative income* hypothesis proposes that the magnitude of the difference between income and that of a reference group is what ultimately matters for health (Deaton, 2001a,b, 2003; Deaton and Paxson, 2004). The average income of the group is mostly used as the point of reference, but other aggregations seem equally plausible, and there is no theoretical guidance on this. The reference group is likely to be unobservable or, at best, observed with error, and this will lead to income inequality reentering the picture even when it exerts no causal effect on individual health (Deaton, 2001b, 2003). The *relative deprivation* hypothesis posits that health is responsive to the difference between income and all larger incomes within the same reference group (Deaton, 2001b; Eibner and Evans, 2005; Yitzhaki, 1979).[56] Lower incomes are assumed to be irrelevant for health, and so the point of reference is individual-specific. The *relative position* hypothesis suggests that the magnitude of income differences is unimportant and that health responds only to rank in the income distribution. This hypothesis is closest to the aforementioned theories that stress the importance of social hierarchy. It might also be used to justify choice of a rank-based measure of income-related health inequality, such as the concentration index (Wagstaff et al., 1991).

17.5.4 Evidence

17.5.4.1 Empirical Challenges

Tests have been performed using data at three levels of aggregation: country, region, and individual. The majority of studies, and most of the early ones, have relied on country-level data, although early US studies used state-level data. Individual-level data is required to discriminate between the five hypotheses (income inequality, absolute income, relative income, relative deprivation, and relative position) because income inequality will

[56] Gravelle and Sutton (2009) also study the opposite situation of individuals caring about being richer than others.

correlate with average population health under all of them (Deaton, 2003; Lynch et al., 2004b; Mackenbach, 2002; Wagstaff and van Doorslaer, 2000).

The relative hypotheses provoke many unanswered questions with respect to the reference groups: how are they formed, does each individual have a unique reference, and how are they to be defined in the data? Testing is further complicated by the potential for position in relation to the reference to be endogenous through choices of group membership. Data quality and the reliability of measures of income inequality are other major issues (Deaton, 2003; West, 1997). Estimates of income inequality or relative income at town or village levels may be derived from relatively few observations and so lack precision (Leigh et al., 2009).

A major problem for analyses identifying effects from cross-country or regional variation is that time-invariant unobservable determinants of health may be correlated with income inequality. Fixed effects methods are unlikely to prove successful at aggregated levels because income inequality tends to evolve rather slowly, and measurement error bias is compounded (Babones, 2008). In addition, fixed effect estimators only identify short run effects and may fail to detect inequality effects operating with a lag. Reverse causality is unlikely to be a major problem at more aggregated levels, but through one or more of the mechanisms identified in Section 17.3, relative income could certainly be a function of health. This would tend to induce bias toward concluding that low relative income exerts a negative impact on health. These limitations must be kept in mind in interpreting the evidence.

We differentiate the evidence by the nature of the hypothesis tested and the level of data aggregation. Given that the impact of income on health is covered in Section 17.4, we do not explicitly consider evidence for the absolute income hypothesis, although we do note what happens to the health–income inequality relationship when individual income is controlled for. Studies published since previous reviews and a few key earlier papers are summarized in Tables 17.8–17.10.

17.5.4.2 Income Inequality Hypothesis

Previous reviews have concluded that the evidence does not point to income inequality as an important determinant of individual health, and this seems to hold for both morbidity and mortality (Deaton, 2003; Leigh et al., 2009; Lynch et al., 2004b; Subramanian and Kawachi, 2004; Wagstaff and van Doorslaer, 2000).[57] They also infer from the literature that state-level income inequality associates negatively with health in the USA, but this is not true in other countries, and this difference most likely reflects racial composition at the state level, although this interpretation is disputed (Subramanian and Kawachi, 2004). There is agreement on the importance of appropriately defining

[57] Wilkinson and Pickett, 2006 dissent, referring mainly to studies using country-level or state-level analyses to infer that income inequality is important for health.

Table 17.8 Cross-country evidence on population health-income inequality association

Authors	Countries/region and period	Data	Estimator	(Partial) correlation with income inequality
Pascual et al. (2005)	12 EU countries 1994–2001	ECHP and OECD	Linear RE and FE	LE −, U5MR +
Cantarero et al. (2005)	12 EU countries 1994–2001	ECHP and OECD	Linear RE and FE	LE −, U5MR +
Babones (2008)	134 countries 1970–1995	WIID and World Bank	OLS, FD	LE −, IMR +, murder rate ns
Biggs et al. (2010)	22 Latin American countries 1960–2007	WDID, GTD, WIID, SEDLAC	Linear FE	LE ns, IMR ns
Wilkinson and Pickett (2010)	25 rich income countries 2000, 2001, 2002, 2003, 2004	UNHDR, WDID, IOT, WHO	Bivariate association	LE −, IMR +, mental health −, obesity +
Regidor et al. (2012)	21 OECD countries 1995, 2000, 2005	OECD	Bivariate association	1995 IMR +, 2005 IMR ns
Avendano (2012)	34 OECD countries 1960–2008	WIID, OECD	Poisson FE	IMR ns
Tacke and Waldmann (2013)	93 countries 1999–2005	WIID, WDID, GHN	OLS	LE −, IMR +, U5MR +

Notes: LE − indicates that life expectancy is negatively correlated with income inequality. IMR + indicates the infant mortality is positively correlated with income inequality. Acronyms of other health indicators, datasets, and estimators are explained in Table A2. ns indicates no significant association.

reference groups: when references are defined with respect to smaller geographic units, such as towns or cities, there is less evidence of an association between income inequality and health.

17.5.4.2.1 Cross-Country Data

The positive cross-country association between mortality and income inequality is well documented and has been confirmed for 12 European countries (Cantarero et al., 2005; Pascual et al., 2005), 25 high-income countries (Wilkinson and Pickett, 2010), and across many countries worldwide (Babones, 2008; Tacke and Waldmann, 2013) (Table 17.8). But the positive association between infant mortality and income inequality across OECD countries evident in 1995 was no longer apparent in 2005 data (Regidor et al., 2012). It also appears that the relationship between population health and income inequality is reversed across low-income countries (Nilsson and Bergh, 2013).

Table 17.9 US cross-region evidence of population health-income inequality association

Authors	Regional unit and period	Data	Estimator	Health measure	(Partial) correlation with income inequality
Ash and Robinson (2009)	287 MSA 1990	CMF, STF	WLS	Age-adjusted mortality ratio	Varies with size of MSA
Deaton and Lubotsky (2009)	287 MSA 1980, 1990	CMF, PUMS	WLS	Age-adjusted mortality ratio	No robust association (conditional on racial composition)
Wilkinson and Pickett (2010)	50 states 1999–2002	Census and CHS, NHANES, BRFSS	Bivariate association	LE, infant deaths, obesity	LE −, IMR +, obesity +
Yang et al. (2012)	3072 counties 1998–2002	CMF	QR	Age-adjusted mortality ratio	+ with effect increasing in magnitude until 80th percentile of mortality

Notes: −/+ indicates negative/positive association of health indicator with income inequality. See Table A2 for explanation of acronyms. MSA is Metropolitan Statistical Area.

The association between population health, measured by life expectancy or infant mortality, and income inequality that is observed across countries is not evident in time variation within countries, even when exploiting very long time series. For example, there is no association between income inequality and mortality between 1900 and 1998 in the USA (Lynch et al., 2004a). Data for Australia, Canada, New Zealand, the USA, and eight European countries display no associations between the within-country evolution of either life expectancy or infant mortality and the income share of the richest 10% (contemporaneous and lagged) between 1903 and 2003 (Leigh and Jencks, 2007). There is also no relationship in the data for 22 Latin American countries between 1960 and 2007 (Biggs et al., 2010). There is an association between the change in income inequality, on the one hand, and the change in life expectancy or infant mortality, on the other hand, for over 90 countries between 1975 and 1995, but this disappears when the change in GDP per capita is controlled for (Babones, 2008). Finally, Avendano (2012) finds no association between within-country variation in infant mortality and income inequality over 4 decades for 34 OECD countries. This finding remains unchanged after allowing for country-specific (linear) time trends or allowing for a lag of 15 years between changes in income inequality and changes in infant mortality.

Table 17.10 Individual-level evidence of association between health and income, relative income, and income inequality

Authors	Country, region, period	Data	Estimator	Findings
Gerdtham and Johannesson (2004)	Sweden, 284 municipalities 1980–1986	ULF, NCD, and NITS register data	Cox proportional hazard	10–17 yr. survival: + AI, ns RI, ns INEQ
Li and Zhu (2006)	China 180 communities 1993	CHNS	Probit	*Excellent/good* SAH: + AI, ns RD, ns RP, ∩ INEQ Physical conditions: ∼AI, ns RD, ns RP, ∼ INEQ (physical conditions = ADL)
Jones and Wildman (2008)	UK 1991–2001	BHPS	OLS, FE, RE	*Good* SAH: + AI, ns RD GHQ: ∼ AI, ∼ RD
Lorgelly and Lindley (2008)	UK 19 regions 1991–2002	BHPS	Pooled, RE, and Mundlak ordered probit	Better SAH: + AI, ns RI, ns INEQ
Petrou and Kupek (2008)	UK 2003	HSE	WLS	EQ-5D: + SC
Gravelle and Sutton (2009)	UK 11 areas 1979–2000	GHS	Pooled binary and ordered probit	Better SAH: + AI, ∼ RI, ∼ INEQ Long-term illness: – AI, ∼ RI; ∼ INEQ
Hildebrand and Kerm (2009)	11 EU countries, 52 EU regions 1994–2001	ECHP	Linear FE	better SAH: ∼ AI, ∼ RI, – INEQ (effect size negligible)
Theodossiou and Zangelidis (2009)	6 EU countries 2004	SOCIOLD	Linear IV	Worse ADL: – AI, + RD Better SAH: + AI, ns RD Mental health: + AI, – RD
d'Hombres et al. (2010)	8 former Soviet countries 2001	LLH	Probit, OLS, GMM	Better SAH: + SC
Karlsson et al. (2010)	21 low-/middle-/high-income countries	FORS, WIID	Pooled ordered probit	Better SAH: + AI, + RI, – INEQ in rich countries (= better ADL)

Table 17.10 Individual-level evidence of association between health and income, relative income, and income inequality—cont'd

Authors	Country, region, period	Data	Estimator	Findings
Mangyo and Park (2011)	China 2004	CIDJ	OLS	Better SAH: + RI, − RD, + RP (= mental health)
van Groezen et al. (2011)	10 EU 2004	SHARE	OLS	Better SAH: + SC
Fang and Rizzo (2012)	China 54 cities and counties 1997–2006	CHNS	FE logit	Better SAH: − INEQ (effect size larger for poorer)
Grönqvist et al. (2012)	Sweden municipalities 1987–2004	Hospital admissions register	Linear FE	Hospital admission: ns INEQ (= sickness leave and mortality)
Lillard et al. (2012)	Australia, Germany, UK, USA	CNEF	Ordered probit	Better SAH: − INEQ
Ronconi et al. (2012)	Argentina 1997	Encuesta de Desarrollo Social	Bivariate probit	Better SAH: + SC
Nilsson and Bergh (2013)	Zambia 155 constituencies, 72 districts, 9 provinces 2004	LCMS IV	OLS and 2SLS	HAZ: + AI, − RI (constituency reference), + RI (provincial reference), + INEQ

Notes: AI, absolute income; RI, relative income; INEQ, income inequality; RD, relative deprivation; RP, relative position; SC, social capital. Read "XXX: AI+" as the health indicator XXX is significantly positively associated with absolute income. Same for RI, INEQ, RD, RP, and SC. Similarly: − indicates negative association; ns indicates no significant association; ~ AI indicates no consistent evidence in favor of or against AI hypothesis (respectively for INEQ, RI, RD, RP, and SC hypotheses); ∩ INEQ indicates an inverse U shape relationship with income inequality. When indicated in the second column, regional unit indicates the level at which income inequality and references for relativities are defined. When not indicated inequality/relativities is at national level. See Table A2 for explanation of other acronyms.

The absence of any evidence that population health moves with changes in income inequality strongly suggests that the static cross-country relationship does not derive from a causal effect of income inequality on health.

17.5.4.2.2 Regional and Cohort Level Data

Conditional on average regional income, life expectancy is sometimes negatively associated with regional income inequality. The negative relationship holds across US states but not at lower levels of aggregation (metropolitan areas, cities)[58] and the evidence is

[58] There is a significant US county level association between income inequality and mortality (Yang et al., 2012).

mixed in other countries (Deaton, 2003; Leigh et al., 2009; Lynch et al., 2004a; Subramanian and Kawachi, 2004; Wagstaff and van Doorslaer, 2000; Wilkinson and Pickett, 2006). The state-level association in the USA could either indicate a causal mechanism or greater aggregation bias at that level. Inclusion of state-level variables eliminates (or dramatically reduces) the association, although these may be mediators rather than confounders (Subramanian and Kawachi, 2004; Wilkinson and Pickett, 2006). The racial composition of US states, a control that knocks out the effect of income inequality, might itself be related to the provision and quality of publicly provided health care (Ash and Robinson, 2009; Deaton and Lubotsky, 2009) (Table 17.9). That state-level income inequality loses much of its explanatory power for mortality after conditioning on measures of social cohesion and interpersonal trust has been interpreted as indicative of a mechanism operating through social capital (Kawachi et al., 1997). Missing from this argument is evidence of a causal effect of income inequality on social capital.

USA and UK birth cohort studies find no association between mortality and income inequality (Deaton and Paxson, 2001, 2004).

17.5.4.2.3 Individual-Level Data

There is little support for an association between *individual* mortality and income inequality conditional on individual income. This holds both in studies exploiting regional variation in income inequality, which risk confounding from regional health effects, and in studies exploiting within-country time variation in income inequality, which risk confounding from time trends (Deaton, 2003; Lynch et al., 2004a,b; Subramanian and Kawachi, 2004; Wagstaff and van Doorslaer, 2000). Morbidity measures, which are more commonly available at the individual level, also have no association with income inequality outside the USA. There is evidence that income inequality at the state level, but again not lower levels, is negatively correlated with the physical and mental health of the poorest individuals in the USA. But this could simply result from state-level income inequality picking up the effect of state-level differences in public policies toward the poor (Mellor and Milyo, 2002) (Table 17.10).

Analyses of individual-level (pseudo) panel data find no association between mortality or morbidity and income inequality. Using high-quality Swedish administrative data with more than 10 years of follow up on income and vital status, Gerdtham and Johannesson (2004) find no effect of municipality-level income inequality on mortality after conditioning on individual income and average municipality income. There is also no relationship of income inequality with SAH revealed by analyses of 12 years of British panel data allowing for unobserved heterogeneity (Lorgelly and Lindley, 2008) and 22 years of British repeated cross-section data (Gravelle and Sutton, 2009). Neither is mental health correlated with income inequality in Australian panel data (Bechtel et al., 2012).

Cross-country studies of individual-level data largely corroborate the negative finding. There are statistically significant, but economically negligible, effects of regional and nationwide income inequality on SAH in panel data on 11 European countries between 1994 and 2001 (Hildebrand and Kerm, 2009). Combining micro data from Australia, Germany, the UK, and the USA with country-level tax records on the income share of the richest percentile (Atkinson et al., 2011), Lillard et al. (2012) find that a higher-income share of the rich is associated with worse SAH, but once time trends are accounted for, the pattern reverses or disappears. The authors also find no evidence that income inequality during the first 20 years of life impacts on current SAH.

Pooling cross-section data for 21 countries, and subject to imposing the same relationship between health and individual income across all those countries, Karlsson et al. (2010) find that health is negatively correlated with income inequality in high-income countries, but there is no relationship in middle- and low-income countries. However, with point-in-time cross-country variation, one can never be sure that the income inequality effect is distinguished from health variation across countries for any other reason. In China, SAH has been found to be positively associated with community-level income inequality at relatively low levels of inequality, but the association turns negative at higher inequality (Li and Zhu, 2006). Correcting for unobserved heterogeneity using panel data, the negative relationship persists and is stronger for poorer individuals (Fang and Rizzo, 2012). While consistent with the income inequality hypothesis, it is possible that this finding is driven by nonlinearity between income and health that is not fully captured in the specification adopted. Child undernutrition in a much lower-income country, Zambia, has been found to be negatively correlated with economic inequality (Nilsson and Bergh, 2013).

If there is health-related migration across regions with differing levels of inequality, perhaps because of differences in medical care, then estimates from the regression of individual health on regional inequality will be biased. One study avoids this by using the random assignment of refugees to a first area of residence in Sweden (Grönqvist et al., 2012). Despite the fact that the range of income inequality across years and municipalities is—perhaps surprisingly—of similar magnitude in Sweden, as it is in the USA or the UK, hospitalizations, sickness leave and mortality were all found to be unrelated to municipality income inequality. Because refugees are likely to be much poorer and to have quite different references from the general population, one may doubt whether this analysis reveals much about either the income inequality or relative hypotheses that can be generalized.

Finally, some studies find that higher levels of social capital are associated with better individual health (Petrou and Kupek, 2008; van Groezen et al., 2011), and others confirm this when using instruments to deal with potential endogeneity of measures of social capital (d'Hombres et al., 2010; Ronconi et al., 2012). None of these studies test whether social capital itself responds to income inequality.

17.5.4.3 Relative Hypotheses

The only study that seeks to test all the relative hypotheses (income, deprivation, and position) rejects all three in favor of the absolute income hypothesis as an explanation of variation in SAH, using UK longitudinal data (Lorgelly and Lindley, 2008). Inconsistent with the relative income hypothesis, conditional on individual income, mortality is found to be lower among individuals living in Swedish municipalities with higher average incomes (Gerdtham and Johannesson, 2004). In contrast, mortality risk, especially for black males, is positively correlated with community average income in the USA (Miller and Paxson, 2006). Data on SAH from 11 European countries are consistent with the relative income hypothesis for males, although the negative correlation of health with average regional income is very small in magnitude, but, if anything, women report slightly better health when regional income is higher (Hildebrand and Kerm, 2009). There is little evidence in favor of the relative income hypothesis in the UK data on SAH (Gravelle and Sutton, 2009).

Analyses of data from Australia (Bechtel et al., 2012), China (Li and Zhu, 2006), and the UK (Gravelle and Sutton, 2009; Jones and Wildman, 2008) find little or no evidence consistent with the relative deprivation hypothesis.[59] Although one of the UK studies finds that mental health does fall slightly with relative deprivation (Jones and Wildman, 2008), which is consistent with other evidence that mental health, but not physical health or longevity, is negatively associated with relative deprivation (Adjaye-Gbewonyo and Kawachi, 2012). One analysis of the US data on mortality and SAH does find evidence consistent with the relative deprivation hypothesis when reference groups are defined narrowly (based on race, state, education, and age) rather than more broadly (state only), as is the case with most other studies (Eibner and Evans, 2005).

17.5.5 Conclusion

The claim that income inequality is harmful to health has provoked much empirical research. This research has delivered little credible evidence to support the hypothesis that income inequality negatively impacts the health of all individuals in society, however. Average population health is negatively associated with income inequality across high- and middle-income countries, but there is no association through time or across regions within countries, except in the USA where state-level differences in health seem to be related to racial composition and possibly also social capital. Individual-level data on morbidity and mortality from high-income countries display no significant, nonnegligible relationship to income inequality after controlling for individual income. Few studies have designs capable of testing the hypotheses that relative income, deprivation,

[59] There is a significant negative relationship in some Australian data, but the magnitude is negligible (Bechtel et al., 2012).

or economic position causally impact health. The evidence that exists shows little support for these hypotheses, except maybe for a negative impact of relative deprivation on mental health. But research has yet to fully separate relative from absolute income effects, investigate the strong possibility that health determines relative economic status, and create an appropriate definition, and thus potential endogeneity, for reference groups.[60]

The lack of evidence might reflect a lack of well-defined theory and, consequently, precision in the way in which empirical analyses relate to hypotheses. Several potential mechanisms, including public provision of goods, social capital, psychosocial mechanisms, and pecuniary externalities, have been proposed, but all lack a precise description of how income inequality and/or relative income impact health. The distinction between the hypotheses is not clear; the relative hypotheses are often claimed to imply an effect of income inequality on the health of all individuals, and it is not clear whether the three relative hypotheses are intended to derive from distinct mechanisms. Psychosocial effects are often loosely cited as the main potential mechanism without specification of how, and for whom, relative economic status provokes stress. Further, it is not clear why priority should be given to relative income, as opposed to some other dimension of socioeconomic position, as a cause of psychosocial stress.

A cheap call for research designs capable of identifying the impact of income inequality on health would be unhelpful. Sound identification of a causal effect of income on health is difficult enough. Obtaining exogenous variation in income inequality is an even more daunting task. Rather than further searching for a significant effect of income inequality on health, a more fruitful research agenda would be to directly investigate the causal mechanisms through which health may be related to income inequality. For example, studies have shown that individual health and social capital are associated, but whether this is the result of causality has not received sufficient attention.

17.6. CONCLUSION

This chapter has examined three propositions: health differences generate income inequality, income differences generate health inequality, and income inequality damages health. Grossly simplifying a host of arguments and a vast body of evidence, our verdicts on these three charges are "guilty," "not proven," and "not guilty," respectively.[61] More cautious assessments of the weight of evidence are provided in the conclusions

[60] Some studies have allowed respondents to define their own reference groups (Karlsson et al., 2010; Mangyo and Park, 2011; Theodossiou and Zangelidis, 2009), although this introduces an obvious endogeneity.

[61] "Not proven" is a verdict available to the courts under Scots law. It is issued when the jury or judge is not convinced of the innocence of the accused but finds the evidence insufficient to prove guilt.

to Sections 17.3–17.5, respectively. Rather than repeat the arguments that lead us to these conclusions, here we restrict attention to their normative and research implications.

Chief among the multitude of mechanisms through which ill-health can impinge on income is the loss of earnings arising from reduced productivity combined with institutional inflexibilities that result in adjustment through employment rather than wages, or marginal changes in work intensity. In high-income countries, ill-health is a major cause of labor-force withdrawal in middle age. On pure efficiency grounds, disability insurance (DI) is called for to weaken the dependence of income on health and thus compress the income distribution. But there is a strong moral hazard effect that makes employment even more sensitive to ill-health. Achieving the optimal balance between income replacement and work incentives is perhaps the greatest challenge for policy that seeks to constrain income inequality arising from ill-health. The task is made even more difficult by increasing economic inequality itself, in the context of which DI can further weaken the labor market attachment of the low-skilled facing deteriorating opportunities. Research needs to move beyond identifying the impact of ill-health on exiting from employment to the design of programs and incentives that can help individuals experiencing health problems remain in work.

Early-life experience might be another major route through which health impacts the distribution of income. Exposure to health risks *in utero* and ill-health in infanthood appear to impact earnings capacity both by interfering with the accumulation of human capital and skills and by triggering illnesses in adulthood that disrupt employment. The currently observed income distribution is, to some degree, the product of health events that occurred during the childhood of the current adult population. This contribution to economic inequality will be particularly strong if, as appears to be the case, disadvantaged children, who would have grown up to be poorer in any case, face greater health risks. Policies directed at childhood circumstances, including those intended to break the link between parental socioeconomic status and health, may not only be preferred normatively in pursuit of the goal of equal opportunity (see Chapter 4), but they might also be favored simply for their effectiveness in influencing the distribution of income among adults. However, much of this line of argument is still supposition. The evidence that childhood health is influenced by economic background and determines adult economic outcomes is persuasive but not yet concrete. Fortunately, the pace of progress in this field makes it unnecessary to call for more research on the contribution of early-life health to economic inequality.

Our "not proven" verdict on the contribution of income (and wealth) to health inequality arises from the potential difficulty in detecting an effect if one did, in fact, exist. At least in high-income countries with near-universal health insurance coverage and in which the burden of disease is mostly chronic, economic circumstances are likely to exert a toll on health, if at all, over a lifetime. The empirical strategies that have been employed, such as fixed effects and instrumenting with transitory financial shocks, are incapable of identifying the long-run effects that may be operating. Finding random permanent shocks to health from which to estimate the health impact on income is easier than stumbling across exogenous events that permanently change income and allow its effect on health to be identified. The empirical task would undoubtedly be more manageable if there was more theory available to identify precise mechanisms through which income (wealth) might plausibly impact health. The lack of theory is understandable. Economists are trained to explain the distribution of income, not health. Forty years after Grossman (1972a) introduced the concept of the health production function, it remains a black box. Although all too often cited to motivate study of the relationship between health and some socioeconomic factor, it is seldom more fully specified to make the mechanism of any effect explicit.

Rather than further identifying a reduced form effect of income or wealth on health, we believe it is more fruitful to focus on plausible inputs to the health production function that can be influenced by economic status. For example, establishing the health effect of damp, squalid housing is more feasible than finding the health effect of the income that affords superior quality housing. This is not merely a call for empirical pragmatism. Provided that redistribution policy is motivated, in part, by (health) specific egalitarianism—and we attribute the extensive involvement of governments in the provision of health insurance and medical care as being motivated not only by the correction of market failures but also by concern for the distribution of health—it might be more efficient to enable poor people to live in less unhealthy conditions, rather than redistributing cash to them. Once basic nutritional needs are satisfied and access to medical care has been divorced from the ability to pay, the path leading from income to health seems a very long one.

If one switches attention from the distribution of health to that of well-being, then the association between income and health may be used to justify greater redistribution of income, even in the absence of any causal effect. Assuming well-being increases with both income and health, the positive correlation between them increases inequality in well-being by more than is implied by the inequality in their marginal distributions (Deaton, 2013). Redistribution of income toward those in worse health would reduce inequality in well-being both by compensating for sickness and, on average, by reaching poorer individuals (Deaton, 2002).[62] According to this argument, redistribution is partially motivated by one dimension of well-being (income) compensating for deficiency in another (health). This is not how health-related income transfers are typically justified. The disabled are paid transfers because their earnings capacity is impaired and/or they have higher costs of living. The transfers are made because ill-health has a causal impact on economic living standards. The ethical argument makes a case for income redistribution to the sick simply because they are sick. Courts awarding damages for injuries

[62] A still more effective redistribution policy might be one that operates through a factor, perhaps education, that exerts a causal impact on both income and health (Deaton, 2002).

irrespective of their consequences for earnings or living costs are consistent with these ethics. But government social policies typically are not. Transfers compensate for financial losses, not reductions in other dimensions of welfare.

With respect to the charge that income inequality threatens health, a case could be made for revising the verdict from "not guilty" to "not proven." It is fundamentally difficult to separate any potential effect of income inequality on the individual's health from that of physical, environmental, social, cultural, or economic determinants of health that operate on the level at which income inequality is measured. Identifying the impact of relative income on health is even more challenging than doing so for absolute income given the added complexity of defining and measuring the reference point. But the limitations are not only empirical. There is a lack of precision in the theoretical arguments as to why economic inequality should impact negatively on health.

The conclusions offered above are based on evidence from high-income countries. In low-income countries, in which a substantial fraction of the population may live close to subsistence and only the economically privileged can afford effective medical care, ill-health is not only an important cause of economic inequality but a consequence of it. But it is the absolute living conditions of the poor, and not their relative deprivation, that takes the toll on health.

ACKNOWLEDGMENTS

We thank the editors, Tony Atkinson and François Bourguignon, for their encouragement and detailed comments on drafts of this chapter. Thanks are also due to Hans van Kippersluis and participants of the NH Handbook of Income Distribution conference, particularly Andrew Clark, for suggestions. Hale Koc and Hao Zhang provided excellent research assistance. We are grateful to various organizations for collecting and providing data (see Appendix). We acknowledge support from the National Institute on Aging, under Grant R01AG037398.

APPENDIX

Table A1 Descriptions and means of variables used in analyses in Section 17.2

		Mean		
	Description	**USA**	**Netherlands**	**China**
Self-assessed health (SAH)[a]				
Poor[a]	1 if self-assessed health is *poor*, 0 otherwise	0.031	0.015	0.066
Moderate/fair	1 if self-assessed health is *moderate/fair*, 0 otherwise	0.123	0.156	0.336
Good	1 if self-assessed health is *good*, 0 otherwise	0.339	0.622	0.474
Very good	1 if self-assessed health is *very good*, 0 otherwise	0.394	0.168	0.124

Table A1 Descriptions and means of variables used in analyses in Section 17.2—cont'd

	Description	Mean USA	Mean Netherlands	Mean China
Excellent	1 if self-assessed health is *excellent*, 0 otherwise	0.112	0.039	
ln(income)[b]	Natural logarithm of equivalent gross household income in national currency	10.142	7.786	9.140
Education[c]				
Low[a]	1 if upper secondary education or less, 0 otherwise	0.271	0.413	0.726
Middle	1 if postsecondary nontertiary education, 0 otherwise	0.341	0.245	0.138
high	1 if tertiary education, 0 otherwise	0.388	0.342	0.136
Gender: male	1 if male, 0 if female	0.403	0.470	0.463
Ethnicity: main group	1 for largest ethnic group [white (USA)/Dutch (NL)/Han (China)], 0 otherwise	0.737	0.879	0.880
Age (years)				
20–29[d]	1 if age is 20–29 years, 0 otherwise	0.183	0.099	0.087
30–39	1 if age is 30–39, 0 otherwise	0.197	0.148	0.191
40–49	1 if age is 40–49, 0 otherwise	0.203	0.188	0.239
50–59	1 if age is 50–59, 0 otherwise	0.217	0.211	0.246
60–69	1 if age is 60–69, 0 otherwise	0.135	0.218	0.148
70+	1 if age is 70+, 0 otherwise	0.066	0.136	0.089
Employment status				
Employed[d]	1 if employed, 0 otherwise	0.564	0.548	0.650
Unemployed	1 if not working and report being unemployed, 0 otherwise	0.107	0.028	0.032
Disabled	1 if not working and report being disabled, 0 otherwise	0.070	0.044	0.005
Retired	1 if not working and report being retired, 0 otherwise	0.134	0.226	0.141
Not working	1 if not working and do not report being unemployed/disabled/retired	0.124	0.154	0.172
Number of observations		5.050	4.137	7.694

[a]In the USA and Dutch surveys, respondents report their health in general as being *excellent*, *very good*, *good*, *fair* (USA)/*moderate* (NL), or *poor*. In the Chinese survey, respondents report their health *relative to others of their own age* as *very good*, *good*, *fair*, or *poor*.

[b]Gross household income is before payment of taxes and social security contributions and after receipt of transfers. Annual income for USA and monthly income for NL and China. Household income equivalized through division by the square root of household size, with the result being assigned to each household member.

[c]Education has been classified using the International Standard Classification of Education (ISCED) for the USA and the Netherlands, with low education referring to ISCED < 4, middle education to ISCED = 4, and higher education to ISCED > 4 (UNESCO Institute for Statistics). For China, low education refers to a primary or junior high school degree, middle education to a senior high school degree, and high education to vocational higher education and university higher education.

[d]Reference category in the least squares and interval regressions in Tables 17.1 and 17.2.

Table A2 Acronyms used in tables

	Name/definition
Datasets	
BHPS	British Household Panel Study
BRFSS	Behavioral risk factor surveillance system
CERRA	Leiden University Center for Research on Retirement and Aging Panel
CHS	US National Centre for Health Statistics
CIDJ	Chinese Inequality and Distributive Justice survey project
CMF	Compressed Mortality File of the National Centre for Health Statistics
CNEF	Cross-national equivalent file
ECHP	European Community Household Panel
FORS	Future of Retirement Survey
FSUH	Financial Survey of Urban Housing
GHN	Globalization-Health Nexus database
GHS	General Household Survey
GSOEP	German Socioeconomic Panel
GTD	WHO Global Tuberculosis Database
HRS	Health and Retirement Study
HSE	Health Survey of England
IOT	International Obesity Taskforce
LCMS	Living Condition Monitoring Study
LLH	Living Conditions, Lifestyle, and Health survey
NCD	Swedish National Cause of Death Statistics
NHANES	National Health and Nutrition Examination Survey
NITS	Swedish National Income Tax Statistics
OECD	OECD Health Data
PSID	Panel Study of Income Dynamics
PUMS	US Census Public Use Micro Sample
RHS	Retirement History Study
SALDRU	South African Labour & Development Research Unit survey
SEDLAC	Socio-Economic Database for Latin America and the Caribbean
SHARE	Survey of Health, Ageing and Retirement in Europe
SOCIOLD	Socioeconomic and occupational effects on the health inequality of the older workforce
STF	US Census Summary Tape File 3C
ULF	Statistics Sweden's Survey of Living Conditions
UNHDR	United Nations Development Report
WDID	World Bank World Development Indicators
WHO	Various databases
WIID	WIDER World Income Inequality Database

Table A2 Acronyms used in tables—cont'd

	Name/definition
Health measures	
ADL	Activities of Daily Living
U5MR	Under-5 mortality rate
GHQ	General health questionnaire (psychological health)
HAZ	Height-for-age z-score
HSCL	Hopkins Symptoms Checklist
IMR	Infant mortality
LE	Life expectancy
Major diagnosis	Cancer, heart disease, lung disease (McClellan, 1998—minor)
Minor diagnosis	Hypertension, diabetes, stroke (McClellan, 1998—major), arthritis, back pain
MR	Mortality rate
SAH	Self-assessed health
SB	Stillbirth rate
WHZ	Weight-for-height z-score
Estimators	
DID	Difference-in-differences
DP	Dynamic programming
FD	First difference
FE	Fixed effects
GMM	Generalized method of moments
GOP	Generalized ordered probit
GPSM	Generalized propensity score matching
IV	Instrumental variables
LPM	Linear probability model
MSM	Method of simulated moments
OLS	Ordinary least squares
QR	Quantile regression
RE	Random effects
SML	Simulated maximum likelihood
2SLS	Two-stage least squares
2SQR	Two-stage quantile regression
WLS	Weighted least squares

DATA SOURCES

We use data from the Chinese Health and Nutrition Survey (CHNS), the Netherlands Longitudinal Internet Studies for the Social Sciences (LISS), and the ALP. We thank the organizations that collected and provided these data.

The LISS panel data were collected by CentERdata (Tilburg University, The Netherlands) through its MESS project funded by the Netherlands Organization for Scientific Research.

RAND ALP are proprietary capabilities of the RAND Corporation. They were developed by RAND with its own funds and with the support of numerous clients and grantors who have commissioned social science and economics research and analysis at RAND.

The National Institute of Nutrition and Food Safety, China Center for Disease Control and Prevention, Carolina Population Center, the University of North Carolina at Chapel Hill, the NIH (R01-HD30880, DK056350, and R01-HD38700) and the Fogarty International Center, all contributed to the collection of the CHNS.

We thank the *Equality Trust Fund* for providing the data to reconstruct the first figure in Section 17.5.

REFERENCES

Acemoglu, D., Angrist, J.D., 2001. Consequences of employment protection? The case of the Americans with Disability Act. J. Polit. Econ. 109 (5), 915–957.

Adams, P., Hurd, M.D., Mcfadden, D., Merrill, A., Ribeiro, T., 2003. Healthy, wealthy, and wise? Tests for direct causal paths between health and socioeconomic status. J. Econom. 112 (1), 3–56.

Adda, J., Chandola, T., Marmot, M., 2003. Socio-economic status and health: causality and pathways. J. Econom. 112 (1), 57–64.

Adda, J., Gaudecker, H., Banks, J., 2009. The impact of income shocks on health: evidence from cohort data. J. Eur. Econ. Assoc. 7 (6), 1361–1399.

Adjaye-Gbewonyo, K., Kawachi, I., 2012. Use of the Yitzhaki Index as a test of relative deprivation for health outcomes: a review of recent literature. Soc. Sci. Med. 75 (1), 129–137.

Agüero, J., Carter, M., Woolard, I., 2009. The Impact of Unconditional Cash Transfers on Nutrition: The South African Child Support Grant. University of Cape Town, Cape Town.

Aigner, D.J., Cain, G.G., 1977. Statistical theories of discrimination in labor markets. Ind. Labor. Relat. Rev. 30 (2), 175–187.

Alesina, A., Baqir, R., Easterly, W., 1999. Public goods and ethnic divisions. Q. J. Econ. 114 (4), 1243–1284.

Alesina, A., Di Tella, R., MacCulloch, R., 2004. Inequality and happiness: are Europeans and Americans different? J. Public Econ. 88 (9–10), 2009–2042.

Allin, S., Stabile, M., 2012. Socioeconomic status and child health: what is the role of health care, health conditions, injuries and maternal health? Health Econ. Policy Law 7 (2), 227.

Almond, D., 2006. Is the 1918 influenza pandemic over? Long-term effects of in utero influenza exposure in the post-1940 U.S. population. J. Polit. Econ. 114 (4), 672–712.

Almond, D., Currie, J., 2011. Killing me softly: the fetal origins hypothesis. J. Econ. Perspect. 25 (3), 153–172.

Almond, D., Currie, J., 2012. Chapter 15—human capital development before age five. In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics. Elsevier, Amsterdam, pp. 1315–1486.

Almond, D., Mazumder, B., 2005. The 1918 influenza pandemic and subsequent health outcomes: an analysis of SIPP data. Am. Econ. Rev. 95, 258–262.
Almond, D., Edlund, L., Palme, M., 2009. Chernobyl's subclinical legacy: prenatal exposure to radioactive fallout and school outcomes in Sweden. Q. J. Econ. 124 (4), 1729–1772.
Almond, D., Hoynes, H.W., Schanzenbach, D.W., 2011. Inside the war on poverty: the impact of food stamps on birth outcomes. Rev. Econ. Stat. 93 (2), 387–403.
Amarante, V., Manacorda, M., Miguel, E., Vigorito, A., 2011. Do Cash Transfers Improve Birth Outcomes? Evidence from Matched Vital Statistics, Social Security and Program Data. CEPR Discussion Paper No. DP8740, CEPR, London.
Apouey, B., Clark, A., 2013. Winning big but feeling no better? The effect of lottery prizes on physical and mental health. Centre for Economic Performance, London, LSE Discussion Paper No. 1228.
Apouey, B., Geoffard, P., 2013. Family income and child health in the UK. J. Health Econ. 32 (4), 715–727.
Arrow, K.J., 1973. The theory of discrimination. In: Ashenfelter, O., Rees, A. (Eds.), Discrimination in Labor Markets. Princeton University Press, Princeton, NJ.
Ash, M., Robinson, D.E., 2009. Inequality, race, and mortality in U.S. cities: a political and econometric review of Deaton and Lubotsky (56:6, 1139–1153, 2003). Soc. Sci. Med. 68 (11), 1909–1913.
Atkinson, A.B., Harrison, A.J., 1978. Distribution of Personal Wealth in Britain, 1923–1972. Cambridge University Press, Cambridge.
Atkinson, A.B., Piketty, T., Saez, E., 2011. Top incomes in the long run of history. J. Econ. Lit. 49 (1), 3–71.
Au, D.W.H., Crossley, T.F., Schellhorn, M., 2005. The effect of health changes and long-term health on the work activity of older Canadians. Health Econ. 14 (10), 999–1018.
Autor, D.H., Duggan, M.G., 2003. The rise in the disability rolls and the decline in unemployment. Q. J. Econ. 118 (1), 157–205.
Autor, D.H., Duggan, M.G., 2006. The growth in the social security disability rolls: a fiscal crisis unfolding. J. Econ. Perspect. 20 (3), 71–96.
Avendano, M., 2012. Correlation or causation? Income inequality and infant mortality in fixed effects models in the period 1960–2008 in 34 OECD countries. Soc. Sci. Med. 75 (4), 754–760.
Babones, S.J., 2008. Income inequality and population health: correlation and causality. Soc. Sci. Med. 66 (7), 1614–1626.
Bago D'Uva, T., Van Doorslaer, E., Lindeboom, M., O'Donnell, O., 2008. Does reporting heterogeneity bias measurement of health disparities? Health Econ. 17 (3), 351–375.
Baird, S., Friedman, J., Schady, N., 2011. Aggregate income shocks and infant mortality in the developing world. Rev. Econ. Stat. 93 (3), 847–856.
Baldwin, M., Johnson, W.G., 1994. Labor market discrimination against men with disabilities. J. Hum. Resour. 29 (1), 1–19.
Banerjee, A.V., Duflo, E., 2011. Poor Economics. Public Affairs, New York.
Banks, J., 2006. Economic capabilities, choices and outcomes at older ages. Fisc. Stud. 27 (3), 281–311.
Barker, D.J.P. (Ed.), 1992. Fetal and Infant Origins of Adult Disease. BMJ Publishing Group, London.
Barker, D.J., 1995. Fetal origins of coronary heart disease. Br. Med. J. 311, 171–174.
Barker, D.J.P., Godfrey, K.M., Gluckman, P.D., Harding, J.E., Owens, J.A., Robinson, J.S., 1993. Fetal nutrition and cardiovascular disease in adult life. Lancet 341 (8850), 938–941.
Barro, R.J., 2003. Determinants of economic growth in a panel of countries. Ann. Econ. Finance 4, 231–274.
Barro, R.J., 2013. Health and economic growth. Ann. Econ. Finance 14 (2), 329–366.
Bazzoli, G., 1985. The early retirement decision: new empirical evidence on the influence of health. J. Hum. Resour. 20, 214–234.
Bechtel, L., Lordan, G., Rao, D.S.P., 2012. Income inequality and mental health? Empirical evidence from Australia. Health Econ. 21, 4–17.
Becker, G., 1957. The Economics of Discrimination. University of Chicago Press, Chicago.
Behrman, J.R., Rosenzweig, M.R., 2004. Returns to birthweight. Rev. Econ. Stat. 86 (2), 586–601.
Bell, D., Heitmueller, A., 2009. The Disability Discrimination Act in the UK: helping or hindering employment among the disabled? J. Health Econ. 28 (2), 465–480.

Ben-Porath, Y., 1967. The production of human capital and the life cycle of earnings. J. Polit. Econ. 75, 352–365.

Berger, M.C., 1983. Labour supply and spouse's health: the effects of illness, disability and mortality. Soc. Sci. Q. 64, 494–509.

Berger, M.C., Fleisher, B.M., 1984. Husband's health and wife's labor supply. J. Health Econ. 3, 63–75.

Bhalotra, S., 2010. Fatal fluctuations? Cyclicality in infant mortality in India. J. Dev. Econ. 93 (1), 7–19.

Biggs, B., King, L., Basu, S., Stuckler, D., 2010. Is wealthier always healthier? The impact of national income level, inequality, and poverty on public health in Latin America. Soc. Sci. Med. 71 (2), 266–273.

Black, D., Kermit, D., Sanders, S., 2002. The impact of economic conditions on participation in disability programs: evidence from the coal boom and bust. Am. Econ. Rev. 92 (1), 27–50.

Black, S.E., Devereux, P.J., Salvanes, K.G., 2007. From the cradle to the labor market? The effect of birth weight on adult outcomes. Q. J. Econ. 122 (1), 409–439.

Blau, D., Gilleskie, D., 2001. The effect of health on employment transitions of older men. Worker Well-being in a Changing Labor Market Research in Labor Economics, vol. 20 JAI Press, Amsterdam, pp. 35–65.

Blau, D.M., Riphahn, R.T., 1999. Labor force transitions of older married couples in Germany. Labour Econ. 6 (2), 229–252.

Bleakley, H., 2007. Disease and development: evidence from hookworm eradication in the south. Q. J. Econ. 122 (1), 73–117.

Bliss, C., Stern, N., 1978. Productivity, wages and nutrition: part I: the theory. J. Dev. Econ. 5 (4), 331–362.

Borghol, N., Suderman, M., Mcardle, W., Racine, A., Hallett, M., Pembrey, M., Hertzman, C., Power, C., Szyf, M., 2012. Associations with early-life socio-economic position in adult DNA methylation. Int. J. Epidemiol. 41 (1), 62–74.

Bound, J., 1991. Self reported versus objective measures of health in retirement models. J. Hum. Resour. 26, 107–137.

Bound, J., Burkhauser, R.V., 1999. Chapter 51: economic analysis of transfer programs targeted on people with disabilities. In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics. Elsevier, Amsterdam, pp. 3417–3528.

Bound, J., Schoenbaum, M., Stinebrickner, T.R., Waidmann, T., 1999. The dynamic effects of health on the labor force transitions of older workers. Labour Econ. 6 (2), 179–202.

Bound, J., Stinebrickner, T., Waidmann, T., 2010. Health, economic resources and the work decisions of older men. J. Econom. 156 (1), 106–129.

Breslau, N., Deldotto, J.E., Brown, G.G., Kumar, S., Ezhuthachan, S., Hufnagle, K.G., Peterson, E.L., 1994. A gradient relationship between low birth weight and IQ at age 6 years. Arch. Pediatr. Adolesc. Med. 148 (4), 377–383.

Brown, S., Roberts, J., Taylor, K., 2010. Reservation wages, labour market participation and health. J. R. Stat. Soc. A Stat. Soc. 173 (3), 501–529.

Burgess, S., Propper, C., Rigg, J.A., 2004. The Impact of Low Income on Child Health: Evidence from a Birth Cohort Study. Centre for Analysis of Social Exclusion, London, LSE CASE Paper 85.

Cai, L., 2009. Effects of health on wages of Australian men. Econ. Rec. 85 (270), 290–306.

Cain, G., 1986. The economic analysis of labor market discrimination: a survey. In: Ashenfelter, O., Layard, R. (Eds.), Handbook of Labor Economics. In: vol. I. North Holland, Amsterdam.

Cantarero, D., Pascual, M., María Sarabia, J., 2005. Effects of income inequality on population health: new evidence from the European community household panel. Appl. Econ. 37 (1), 87–91.

Carman, K.G., 2013. Inheritances, intergenerational transfers, and the accumulation of health. Am. Econ. Rev. Papers & Proc. 103 (3), 451–455.

Carter, H., Glick, P., 1976. Marriage and Divorce: A Social and Economic Study. Harvard University Press, Cambridge, MA.

Case, A., 2004. Does money protect health status? Evidence from South African pensions. Perspectives on the Economics of Aging, University of Chicago Press, Chicago, IL, pp. 287–312.

Case, A., Paxson, C., 2008. Stature and status: height, ability, and labor market outcomes. J. Polit. Econ. 116 (3), 499–532.
Case, A., Paxson, C., 2011. The long reach of childhood health and circumstance: evidence from the Whitehall II study. Econ. J. 121 (554), F183–F204.
Case, A., Lubotsky, D., Paxson, C., 2002. Economic status and health in childhood: the origins of the gradient. Am. Econ. Rev. 92 (5), 1308–1334.
Case, A., Fertig, A., Paxson, C., 2005. The lasting impact of childhood health and circumstance. J. Health Econ. 24, 365–389.
Case, A., Lee, D., Paxson, C., 2008. The income gradient in children's health: a comment on Currie, Shields, and Wheatley Price. J. Health Econ. 27 (3), 801–807.
Cawley, J., Ruhm, C.J., 2011. The economics of risky health behaviors 1. Handb. Health Econom., vol. 2, p. 95.
Cawley, J., Heckman, J.J., Vytlacil, E.J., 2001. Three observations on wages and measured cognitive ability. Labour Econ. 8 (4), 419–442.
Cawley, J., Moran, J., Simon, K., 2010. The impact of income on the weight of elderly Americans. Health Econ. 19 (8), 979.
Chandola, T., Bartley, M., Sacker, A., Jenkinson, C., Marmot, M., 2003. Health selection in the Whitehall II study. Soc. Sci. Med. 56 (10), 2059–2072.
Chang, F., 1991. Uncertain lifetimes, retirement and economic welfare. Economica 58 (230), 215–232.
Charles, K.K., 1999. Sickness in the Family: Health Shocks and Spousal Labor Supply. University of Michigan, Ann Arbor.
Charles, K.K., 2003. The longitudinal structure of earnings losses among work-limited disabled workers. J. Hum. Resour. 38 (3), 618–646.
Clark, A.E., Frijters, P., Shields, M.A., 2008. Relative income, happiness, and utility: an explanation for the Easterlin paradox and other puzzles. J. Econ. Lit. 46 (1), 95–144.
Coile, C.C., 2004. Health Shocks and Couples' Labor Supply Decisions. NBER Working Paper No. 10810, NBER, Cambridge, MA.
Commission on Macroeconomics and Health, 2001. Macroeconomics and Health: Investing in Health for Economic Development. World Health Organisation, Geneva.
Commission on the Social Determinants of Health, 2008. Closing the gap in a generation: health equity through action on the social determinants of health. Final Report of the Commission on Social Determinants of Health,World Health Organization, Geneva.
Condliffe, S., Link, C.R., 2008. The relationship between economic status and child health: evidence from the United States. Am. Econ. Rev. 98 (4), 1605–1618.
Conley, D., Bennett, N.G., 2001. Birth weight and income: interactions across generations. J. Health Soc. Behav. 42 (4), 450–465.
Conti, G., Heckman, J.J., Urzua, S., 2010a. Early endowments, education, and health. Unpublished manuscript, University of Chicago, Department of Economics, Chicago, p. 162.
Conti, G., Heckman, J., Urzua, S., 2010b. The education–health gradient. Am. Econ. Rev. 100, 234–238, Papers and proceedings.
Contoyannis, P., Rice, N., 2001. The impact of health on wages: evidence from the British Household Panel Survey. Emp. Econ. 26 (4), 599–622.
Contoyannis, P., Jones, A.M., Rice, N., 2004. The dynamics of health in the British Household Panel Survey. J. Appl. Econ. 19 (4), 473–503.
Costello, E.J., Worthman, C., Erkanli, A., Angold, A., 2007. Prediction from low birth weight to female adolescent depression: a test of competing hypotheses. Arch. Gen. Psychiatry 64 (3), 338–344.
Crimmins, E.M., Beltran-Sanchez, H., 2011. Mortality and morbidity trends: is there compression of morbidity? J. Gerentol. B 66B (1), 75–86.
Cunha, F., Heckman, J., 2007. The technology of skill formation. Am. Econ. Rev. 97 (2), 31–47.
Cunha, F., Heckman, J., 2008. A new framework for the analysis of inequality. Macroecon. Dyn. 12 (Suppl. S2), 315.
Cunha, F., Heckman, J.J., 2009. The economics and psychology of inequality and human development. J. Eur. Econ. Assoc. 7 (2–3), 320–364.

Cunha, F., Heckman, J.J., Lochner, L., Masterov, D.V., 2006. Chapter 12: interpreting the evidence on life cycle skill formation. In: Hanushek, E., Welch, F. (Eds.), Handbook of the Economics of Education. Elsevier, Amsterdam, pp. 697–812.

Currie, J., 2009. Healthy, wealthy, and wise: socioeconomic status, poor health in childhood, and human capital development. J. Econ. Lit. 47 (1), 87–122.

Currie, J., Hyson, R., 1999. Is the impact of health shocks cushioned by socioeconomic status? The case of low birthweight. Am. Econ. Rev. 89 (2), 245–250.

Currie, J., Madrian, B.C., 1999. Chapter 50: health, health insurance and the labor market. In: Ashenfelter, O.C., Card, D. (Eds.), Handbook of Labor Economics. Elsevier, Amsterdam, pp. 3309–3416.

Currie, J., Moretti, E., 2007. Biology as destiny? Short- and long-run determinants of intergenerational transmission of birth weight. J. Labor Econ. 25 (2), 231–264.

Currie, J., Stabile, M., 2003. Socioeconomic status and child health: why is the relationship stronger for older children. Am. Econ. Rev. 93 (5), 1813–1823.

Currie, J., Stabile, M., 2006. Child mental health and human capital accumulation: the case of ADHD. J. Health Econ. 25 (6), 1094–1118.

Currie, A., Shields, M.A., Price, S.W., 2007. The child health/family income gradient: evidence from England. J. Health Econ. 26 (2), 213–232.

Currie, J., Stabile, M., Manivong, P., Roos, L.L., 2010. Child health and young adult outcomes. J. Hum. Resour. 45 (3), 517–548.

Cutler, D.M., Lleras-Muney, A., 2010. Understanding differences in health behaviors by education. J. Health Econ. 29 (1), 1–28.

Cutler, D.M., Knaul, F., Lozano, R., Méndez, O., Zurita, B., 2002. Financial crisis, health outcomes and ageing: Mexico in the 1980s and 1990s. J. Public Econ. 84 (2), 279–303.

Cutler, D.M., Lange, F., Meara, E., Richards-Shubik, S., Ruhm, C.J., 2011a. Rising educational gradients in mortality: the role of behavioral factors. J. Health Econ. 30 (6), 1174–1187.

Cutler, D., Lleras-Muney, A., Vogl, T., 2011b. Chapter 7: socioeconomic status and health: dimensions and mechanisms. In: Glied, S., Smith, P.C. (Eds.), Oxford Handbook of Health Economics. Oxford University Press, Oxford, pp. 124–163.

Cutler, D., Ghosh, K., Landrum, M.B., 2013. Evidence for significant compression of morbidity in the elderly US population. NBER Working Paper No. 19268, NBER, Cambridge, MA.

Daly, M.C., Bound, J., 1996. Worker adaptation and employer accommodation following the onset of a health impairment. J. Gerontol. 51, S53–S60.

Dasgupta, P., 1993. An Inquiry into Well-Being and Destitution. Oxford University Press, Oxford.

Dasgupta, P., 1997. Nutritional status, the capacity for work, and poverty traps. J. Econom. 77 (1), 5–37.

Dasgupta, P., Ray, D., 1986. Inequality as a determinant of malnutrition and unemployment: theory. Econ. J. 96 (384), 1011–1034.

Deaton, A., 2001a. Inequalities in income and inequalities in health. In: Welch, F. (Ed.), The Causes and Consequences of Increasing Inequality. Chicago University Press, Chicago, pp. 285–313.

Deaton, A., 2001b. Relative Deprivation, Inequality, and Mortality. National Bureau of Economic Research, Cambridge, MA.

Deaton, A., 2002. Policy implications of the gradient of health and wealth. Health Aff. 21 (2), 13–30.

Deaton, A., 2003. Health, inequality, and economic development. J. Econ. Lit. 41 (1), 113–158.

Deaton, A., 2013. What does the empirical evidence on SES and health tell us about inequity and about policy? In: Eyal, N., Hurst, S.A., Norheim, O.F., Wikler, D. (Eds.), Inequalities in Health: Concepts, Measures and Ethics. Oxford University Press, New York.

Deaton, A., Lubotsky, D., 2009. Income inequality and mortality in U.S. cities: weighing the evidence. A response to ash. Soc. Sci. Med. 68 (11), 1914–1917.

Deaton, A., Paxson, C., 2001. Mortality, education, income, and inequality among American cohorts. In: Wise, D. (Ed.), Themes in the Economics of Aging. University of Chicago Press, Chicago, IL, pp. 129–170.

Deaton, A., Paxson, C., 2004. Mortality, income, and income inequality over time in Britain and the United States. In: Wise, D. (Ed.), Perspectives on the Economics of Ageing. University of Chicago Press, Chicago, IL, pp. 247–285.
Deleire, T., 2000. The wage and employment effects of the Americans with Disabilities Act. J. Hum. Resour. 35 (4), 693–715.
Deleire, T., 2001. Changes in wage discrimination against people with disabilities: 1984–93. J. Hum. Resour. 36 (1), 144–158.
Deolalikar, A.B., 1988. Nutrition and labor productivity in agriculture: estimates for rural South India. Rev. Econ. Stat. 70 (3), 406–413.
D'Hombres, B., Rocco, L., Suhrcke, M., McKee, M., 2010. Does social capital determine health? Evidence from eight transition countries. Health Econ. 19 (1), 56–74.
Diamond, P., Sheshinski, E., 1995. Economic aspects of optimal disability benefits. J. Public Econ. 57 (1), 1–23.
Disney, R., Emmerson, C., Wakefield, M., 2006. Ill health and retirement in Britain: a panel data-based analysis. J. Health Econ. 25 (4), 621–649.
Dobkin, C., Puller, S., 2007. The effects of government transfers on monthly cycles in drug abuse, hospitalization and mortality. J. Pub. Econ. 91, 2137–2151.
Doorslaer, E.V., Jones, A.M., 2003. Inequalities in self-reported health: validation of a new approach to measurement. J. Health Econ. 22 (1), 61–87.
Duflo, E., 2000. Child health and household resources in South Africa: evidence from the old age pension program. Am. Econ. Rev. 90 (2), 393–398.
Duflo, E., 2003. Grandmothers and granddaughters: old-age pensions and intrahousehold allocation in South Africa. World Bank Econ. Rev. 17 (1), 1–25.
Dwyer, D.S., Mitchell, O.S., 1999. Health problems as determinants of retirement: are self-rated measures endogenous? J. Health Econ. 18 (2), 173–193.
Ebrahim, S., 2012. Epigenetics: the next big thing. Int. J. Epidemiol. 41 (1), 1–3.
Eibner, C., Evans, W.N., 2005. Relative deprivation, poor health habits, and mortality. J. Hum. Resour. XL (3), 591–620.
Erreygers, G., 2009. Can a single indicator measure both attainment and shortfall inequality? J. Health Econ. 28 (4), 885–893.
Erreygers, G., Van Ourti, T., 2011. Measuring socioeconomic inequality in health, health care and health financing by means of rank-dependent indices: a recipe for good practice. J. Health Econ. 30 (4), 685–694.
Ettner, S.L., 1996. New evidence on the relationship between income and health. J. Health Econ. 15 (1), 67–85.
Evans, W.N., Moore, T.J., 2011. The short-term mortality consequences of income receipt. J. Public Econ. 95 (11), 1410–1424.
Fang, H., Rizzo, J.A., 2012. Does inequality in China affect health differently in high-versus low-income households? Appl. Econ. 44 (9), 1081–1090.
Feeny, D., Furlong, W., Torrance, G.W., Goldsmith, C.H., Zhu, Z., Depauw, S., Denton, M., Boyle, M., 2002. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. Med. Care 40 (2), 113–128.
Fernald, L.C., Hidrobo, M., 2011. Effect of Ecuador's cash transfer program (*Bono de Desarrollo Humano*) on child development in infants and toddlers: a randomized effectiveness trial. Soc. Sci. Med. 72 (9), 1437–1446.
Fernald, L.C., Gertler, P.J., Neufeld, L.M., 2008. Role of cash in conditional cash transfer programmes for child health, growth, and development: an analysis of Mexico's Oportunidades. Lancet 371 (9615), 828–837.
Finkelstein, A., Luttmer, E.F.P., Notowidigdo, M.J., 2013. What good is wealth without health? The effect of health on the marginal utility of consumption. J. Eur. Econ. Assoc. 11 (s1), 221–258.
Fleurbaey, M., Schokkaert, E., 2011. Chapter Sixteen—equity in health and health care. In: Pauly, M.V., Mcguire, T.G., Barros, P.P. (Eds.), Handbook of Health Economics. In: vol. 2. Elsevier, Amsterdam, pp. 1003–1092.
French, E., 2005. The effects of health, wealth, and wages on labour supply and retirement behaviour. Rev. Econ. Stud. 72 (2), 395–427.

Frijters, P., Shields, M.A., Haisken-DeNew, J.P., 2005. The effect of income on health: evidence from a large scale natural experiment. J. Health Econ. 24, 997–1017.
Fu, H., Goldman, N., 1994. Are healthier people more likely to marry? An event history analysis based on the NLSY. Working Paper 94-5, Office of Population Research, Princeton University, New Jersey.
Galama, T.J., Van Kippersluis, H., 2010. A Theory of Socioeconomic Disparities in Health over the Life Cycle. Tinbergen Institute, Amsterdam, Discussion Paper No. 2010-08-24.
García Gómez, P., Lopez-Nicolas, A., 2006. Health shocks, employment and income in the Spanish labour market. Health Econ. 9, 997–1009.
García Gómez, P., Van Kippersluis, H., O'Donnell, O., Van Doorslaer, E., 2013. Long term and spillover effects of health on employment and income. J. Hum. Resour. 48 (4), 873–909.
García-Gómez, P., 2011. Institutions, health shocks and labour market outcomes across Europe. J. Health Econ. 30 (1), 200–213.
Gardner, J., Oswald, A.J., 2007. Money and mental wellbeing: a longitudinal study of medium-sized lottery wins. J. Health Econ. 26 (1), 49–60.
Gerdtham, U., Johannesson, M., 2004. Absolute income, relative income, income inequality, and mortality. J. Hum. Resour. XXXIX (1), 228–247.
Gertler, P., 2004. Do conditional cash transfers improve child health? Evidence from PROGRESA's control randomized experiment. Am. Econ. Rev. 94 (2), 336–341.
Gertler, P., Gruber, J., 2001. Insuring consumption against illness. Am. Econ. Rev. 92 (1), 51–70.
Goda, G.S., Golberstein, E., Grabowski, D.C., 2011. Income and the utilization of long-term care services: evidence from the social security benefit notch. J. Health Econ. 30 (4), 719–729.
Goldman, N., 1993. Marriage selection and mortality patterns: inferences and fallacies. Demography 30 (2), 189–208.
Granger, C.W.J., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37 (3), 424–438.
Grantham-Mcgregor, S.M., Powell, C.A., Walker, S.P., Himes, J.H., 1991. Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: the Jamaican Study. Lancet 338 (8758), 1–5.
Grantham-Mcgregor, S., Cheung, Y.B., Cueto, S., Glewwe, P., Richter, L., Strupp, B., 2007. Developmental potential in the first 5 years for children in developing countries. Lancet 369, 60–70.
Gravelle, H., 1998. How much of the relation between population mortality and unequal distribution of income is a statistical artefact? Br. Med. J. 316, 382–385.
Gravelle, H., Sutton, M., 2009. Income, relative income, and self-reported health in Britain 1979–2000. Health Econ. 18 (2), 125–145.
Gravelle, H., Wildman, J., Sutton, M., 2002. Income, income inequality and health: what can we learn from aggregate data? Soc. Sci. Med. 54 (4), 577–589.
Grönqvist, H., Johansson, P., Niknami, S., 2012. Income inequality and health: lessons from a refugee residential assignment program. J. Health Econ. 31 (4), 617–629.
Grossman, M., 1972a. The Demand for Health: A Theoretical and Empirical Investigation. NBER, New York.
Grossman, M., 1972b. On the concept of health capital and the demand for health. J. Polit. Econ. 80 (2), 223–255.
Grossman, M., 2000. The human capital model. In: Culyer, A., Newhouse, J. (Eds.), North Holland Handbook in Health Economics. North Holland, Amsterdam, Netherlands, pp. 1804–1862.
Grossman, M., Kaestner, R., 1997. Effects of education on health. In: Behrman, J.R., Stacey, N. (Eds.), The Social Benefits of Education. University of Michigan Press, Ann Arbor, pp. 69–123.
Gruber, J., 2000. Disability insurance benefits and labor supply. J. Polit. Econ. 108 (6), 1162–1183.
Gustman, A.L., Steinmeier, T.L., 1986a. A disaggregated, structural analysis of retirement by race, difficulty of work and health. Rev. Econ. Stat. 68 (3), 509–513.
Gustman, A.L., Steinmeier, T.L., 1986b. A structural retirement model. Econometrica 54 (3), 555–584.
Halla, M., Zweimuller, M., 2013. The effect of health on income: quasi-experimental evidence from commuting accidents. Labour Econ. 24, 23–38.
Hammermesh, D., 1984. Life-cycle effects on consumption and retirement. J. Labor Econ. 2 (3), 353–370.

Hausman, J., Taylor, W., 1981. Panel data and unobservable individual effects. Econometrica 49, 1377–1398.
Hayes, B., Sharif, F., 2009. Behavioural and emotional outcome of very low birth weight infants—literature review. J. Matern. Fetal Neonatal Med. 22 (10), 849–856.
Heckman, J., 2003. Conditioning, causality and policy analysis. J. Econom. 112 (1), 73–78.
Heckman, J.J., 2007. The economics, technology, and neuroscience of human capability formation. Proc. Natl. Acad. Sci. U.S.A. 104 (33), 13250–13255.
Heckman, J.J., Stixrud, J., Urzua, S., 2006. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. J. Labor Econ. 24 (3), 411–482.
Hediger, M.L., Overpeck, M.D., Ruan, W.J., Troendle, J.F., 2002. Birthweight and gestational age effects on motor and social development. Paediatr. Perinat. Epidemiol. 16, 33–46.
Herrnstein, R.J., Murray, C.A., 1994. The Bell Curve: Intelligence and Class Structure in American Life. Free Press, New York.
Hertzman, C., Power, C., Matthews, S., 2001. Using an interactive framework of society and lifecourse to explain self-rated health in early adulthood. Soc. Sci. Med. 53, 1575–1585.
Hibbs, H., 1915. The influence of economic and industrial conditions on infant mortality. Q. J. Econ. 30, 127–151.
Hildebrand, V., Kerm, P.V., 2009. Income inequality and self-rated health status: evidence from the European community household panel. Demography 46 (4), 805–825.
Hille, E.T.M., Ouden, A.L.D., Bauer, L., Oudenrijn, C.V.D., Brand, R., Verloove-Vanhorick, S.P., 1994. School performance at nine years of age in very premature and very low birth weight infants: perinatal risk factors and predictors at five years of age. J. Pediatr. 125 (3), 426–434.
Hotchkiss, J.L., 2004. A closer look at the employment impact of the Americans with Disabilities Act. J. Hum. Resour. 39 (4), 887–911.
Hoynes, H.W., Schanzenbach, D.W., 2009. Consumption responses to in-kind transfers: evidence from the introduction of the food stamp program. Am. Econ. J. Appl. Econ. 1 (4), 109–139.
Hoynes, H., Page, M., Stevens, A.H., 2011. Can targeted transfers improve birth outcomes? Evidence from the introduction of the WIC program. J. Public Econ. 95 (7), 813–827.
Hoynes, H.W., Miller, D.L., Simon, D., 2012. Income, the Earned Income Tax Credit, and Infant Health. NBER, Cambridge, MA, Working Paper No. 18206.
Hugonnier, J., Pelgrin, F., St-Amour, P., 2013. Health and (other) asset holdings. Rev. Econ. Stud. 80, 663–710.
Hurd, M., Kapteyn, A., 2003. Health, wealth, and the role of institutions. J. Hum. Resour. 38 (2), 386–415.
Jackle, R., Himmler, O., 2010. Health and wages. J. Hum. Resour. 45 (2), 364–406.
Jencks, C., 1979. Who Gets Ahead? The Determinants of Economic Success in America. Basic Books, New York.
Jensen, R.T., Richter, K., 2004. The health implications of social security failure: evidence from the Russian pension crisis. J. Public Econ. 88 (1), 209–236.
Johnson, W.G., Lambrinos, J., 1985. Wage discrimination against handicapped men and women. J. Hum. Resour. 20 (2), 264–277.
Johnson, R., Schoeni, R., 2011. The influence of early-life events on human capital, health status, and labor market outcomes over the life course. BE J. Econ. Anal. Policy 11 (3), 1–55.
Jones, A.M., Wildman, J., 2008. Health, income and relative deprivation: evidence from the BHPS. J. Health Econ. 27 (2), 308–324.
Jones, M.K., Latreille, P.L., Sloane, P.J., 2006. Disability, gender, and the British labour market. Oxf. Econ. Papers 58 (3), 407–449.
Jones, A.M., Rice, N., Roberts, J., 2010. Sick of work or too sick to work? Evidence on self-reported health shocks and early retirement from the BHPS. Econ. Model. 27 (4), 866–880.
Kahn, L., 1991. Customer discrimination and affirmative action. Econ. Inq. 29, 555–571.
Kalemli-Ozcan, S., Weil, D., 2010. Mortality change, the uncertainty effect and retirement. J. Econ. Growth 15 (1), 65–91.
Karlsson, M., Nilsson, T., Lyttkens, C.H., Leeson, G., 2010. Income inequality and health: importance of a cross-country perspective. Soc. Sci. Med. 70 (6), 875–885.

Kawachi, I., Kennedy, B.P., 1997. Health and social cohesion: why care about income inequality? Br. Med. J. 314 (7086), 1037–1040.
Kawachi, I., Kennedy, B.P., Lochner, K., Prothrow-Stith, D., 1997. Social capital, income inequality and mortality. Am. J. Public Health 87 (9), 1491–1498.
Kelly, E., 2011. The scourge of Asian flu: in utero exposure to pandemic influenza and the development of a cohort of British children. J. Hum. Resour. 21 (4), 669–694.
Kennedy, B., Kawachi, I., Glass, R., Prothrow-Smith, D., 1998. Income distribution, socioeconomic status, and self rated health in the United States: multilevel analysis. Br. Med. J. 317, 917–921.
Kerkhofs, M., Lindeboom, M., Theeuwes, J., 1999. Retirement, financial incentives and health. Labour Econ. 6 (2), 203–227.
Khanam, R., Nghiem, H.S., Connelly, L.B., 2009. Child health and the income gradient: evidence from Australia. J. Health Econ. 28 (4), 805–817.
Kidd, M.P., Sloane, P.J., Ferko, I., 2000. Disability and the labour market: an analysis of British males. J. Health Econ. 19 (6), 961–981.
Kim, B., Ruhm, C.J., 2012. Inheritances, health and death. Health Econ. 21 (2), 127–144.
Kirkegaard, I., Obel, C., Hedegaard, M., Henriksen, T.B., 2006. Gestational age and birth weight in relation to school performance of 10-year-old children: a follow-up study of children born after 32 completed weeks. Pediatrics 118 (4), 1600–1606.
Kuh, D., Ben-Shlomo, Y. (Eds.), 1997. A Lifecourse Approach to Chronic Disease Epidemiology: Tracing the Origins of Ill-Health from Early to Adult Life. Oxford University Press, Oxford.
Kuh, D.J.L., Wadsworth, M.E.J., 1993. Physical health status at 36 years in a British national birth cohort. Soc. Sci. Med. 37 (7), 905–916.
Kuh, D., Ben-Shlomo, Y., et al., 2003. Life course epidemiology. J. Epidemiol. Commun. Health 57 (10), 778–783.
Lambert, P., Zheng, B., 2011. On the consistent measurement of attainment and shortfall inequality. J. Health Econ. 30 (1), 214–219.
Lawlor, D.A., Clark, H., Davey Smith, G., Leon, D.A., 2006. Intrauterine growth and intelligence within sibling pairs: findings from the Aberdeen children of the 1950s cohort. Pediatrics 117 (5), E894–E902.
Leibenstein, H., 1957. The theory of underemployment in backward economies. J. Polit. Econ. 65 (2), 91–103.
Leigh, A., Jencks, C., 2007. Inequality and mortality: long-run evidence from a panel of countries. J. Health Econ. 26 (1), 1–24.
Leigh, A., Jencks, C., Smeeding, T., 2009. Health and economic inequality. In: Salverda, W., Nolan, B., Smeeding, T. (Eds.), Oxford Handbook of Economic Inequality. Oxford University Press, Oxford, pp. 384–405.
Li, H., Zhu, Y., 2006. Income, income inequality, and health: evidence from China. J. Comp. Econ. 34 (4), 668–693.
Lillard, L., Panis, C.A., 1996. Marital status and mortality: the role of health. Demography 33, 313–327.
Lillard, D.R., Burkhauser, R.V., Hahn, M.H., Wilkins, R., 2012. Does the income share of the top 1 percent predict self-reported health status of the 99 percent? Evidence from four countries. In: 10th International German Socio-Economic Panel User Conference 2012.
Lin, M., Liu, J., 2009. Do lower birth weight babies have lower grades? Twin fixed effect and instrumental variable evidence from Taiwan. Soc. Sci. Med. 68 (10), 1780–1787.
Lindahl, M., 2005. Estimating the effect of income on health and mortality using lottery prizes as an exogenous source of variation in income. J. Hum. Resour. 40 (1), 144–168.
Lindeboom, M., 2012. Health and work of older workers. In: Jones, A.M. (Ed.), The Elgar Companion to Health Economics, second ed. Edward Elgar, Cheltenham, pp. 26–35.
Lindeboom, M., Kerkhofs, M., 2009. Health and work of the elderly: subjective health measures, reporting errors and endogeneity in the relationship between health and work. J. Appl. Econom. 24 (6), 1024–1046.
Loe, I.M., Lee, E.S., Luna, B., Feldman, H.M., 2011. Behavior problems of 9–16 year old preterm children: biological, sociodemographic, and intellectual contributions. Early Hum. Dev. 87, 247–252.

Lorgelly, P.K., Lindley, J., 2008. What is the relationship between income inequality and health? Evidence from the BHPS. Health Econ. 17 (2), 249–265.
Luo, Y., Waite, L.J., 2005. The impact of childhood and adult SES on physical, mental, and cognitive well-being on later life. J. Gerontol. 60B (2), S93–S101.
Lynch, J., Smith, G.D., Harper, S., Hillemeier, M., 2004a. Is income inequality a determinant of population health? Part 2. U.S. national and regional trends in income inequality and age- and cause-specific mortality. Milbank Q. 82 (2), 355–400.
Lynch, J., Smith, G.D., Harper, S., Hillemeier, M., Ross, N., Kaplan, G.A., Wolfson, M., 2004b. Is income inequality a determinant of population health? Part 1. A systematic review. Milbank Q. 82 (1), 5–99.
Mackenbach, J.P., 2002. Income inequality and population health. BMJ 324 (7328), 1–2.
Mangyo, E., Park, A., 2011. Relative deprivation and health. J. Hum. Resour. 46 (3), 459–481.
Marmot, M., Rose, G., Shipley, M., Hamilton, P., 1978. Employment grade and coronary heart disease in British civil servants. J. Epidemiol. Commun. Health 32, 244–249.
Marmot, M., Davey Smith, G., Stansfeld, S., Patel, C., North, F., Head, J., White, I., Brunner, E.J., Feeney, A., 1991. Health inequalities among British civil servants: the Whitehall II study. Lancet 337, 1387–1393.
Mazumder, B., Almond, D., Park, K., Crimmins, E.M., Finch, C.E., 2010. Lingering prenatal effects of the 1918 influenza pandemic on cardiovascular disease. J. Dev. Orig. Health Dis. 1 (1), 26.
McClellan, M., 1998. Health events, health insurance, and labor supply: evidence from the Health & Retirement Study. In: Wise, D. (Ed.), Frontiers in the Economics of Aging. University of Chicago Press, Chicago, IL.
Mcguinness, D., Mcglynn, L.M., Johnson, P.C., Macintyre, A., Batty, G.D., Burns, H., Cavanagh, J., Deans, K.A., Ford, I., Mcconnachie, A., Mcginty, A., Mclean, J.S., Millar, K., Packard, C.J., Sattar, N.A., Tannahill, C., Velupillai, Y.N., Shiels, P.G., 2012. Socio-economic status is associated with epigenetic differences in the pSoBid cohort. Int. J. Epidemiol. 41 (1), 151–160.
Mcinerney, M., Mellor, J.M., 2012. Recessions and seniors' health, health behaviors, and healthcare use: analysis of the Medicare Current Beneficiary Survey. J. Health Econ. 31 (5), 744–751.
Mcinerney, M., Mellor, J.M., Hersch Nicholas, L., 2013. Recession depression: mental health effects of the 2008 stock market crash. J. Health Econ. 32 (6), 1090–1104.
Meer, J., Miller, D.L., Rosen, H.S., 2003. Exploring the health–wealth nexus. J. Health Econ. 22 (5), 713–730.
Mellor, J.M., Milyo, J., 2002. Income inequality and health status in the United States: evidence from the Current Population Survey. J. Hum. Resour. 37 (3), 510–539.
Meltzer, A.H., Richard, S.F., 1981. A rational theory of the size of government. J. Polit. Econ. 89 (5), 914–927.
Michaud, P., Van Soest, A., 2008. Health and wealth of elderly couples: causality tests using dynamic panel data models. J. Health Econ. 27 (5), 1312–1325.
Miguel, E., Kremer, M., 2004. Worms: identifying impacts on education and health in the presence of treatment externalities. Econometrica 72 (1), 159–217.
Miller, D.L., Paxson, C., 2006. Relative income, race, and mortality. J. Health Econ. 25 (5), 979–1003.
Miller, G., Urdinola, B.P., 2010. Cyclicality, mortality, and the value of time: the case of coffee price fluctuations and child survival in Colombia. J. Polit. Econ. 118 (1), 113.
Milligan, K.S., Wise, D.A., 2012. Introduction and summary. In: Wise, D.A. (Ed.), Social Security Programs and Retirement Around the World: Historical Trends in Mortality and Health, Employment, and Disability Insurance Participation and Reforms. The University of Chicago Press, Chicago.
Mirlees, J.A., 1975. Pure theory of underdeveloped economies. In: Reynolds, L.G. (Ed.), Agriculture in Development Theory. Yale University Press, New Haven, CT.
Mocan, N., Raschke, C., Unel, B., 2013. The Impact of Mothers' Earnings on Health Inputs and Infant Health. NBER, Cambridge, MA, Working Paper 19434.
Mohanan, M., 2013. Causal effects of health shocks on consumption and debt: quasi-experimental evidence from bus accident injuries. Rev. Econ. Stat. 95 (2), 673–681.
Moller Dano, A., 2005. Road injuries and long-run effects on income and employment. Health Econ. 14 (9), 955–970.

Moran, J.R., Simon, K.I., 2006. Income and the use of prescription drugs by the elderly evidence from the notch cohorts. J. Hum. Resour. 41 (2), 411–432.
National Bureau of Statistics of China, 2007. China Statistical Yearbook 2007. China Statistics Press, Beijing. http://www.stats.gov.cn/tjsj/ndsj/2007/indexeh.htm.
Navarro, V., 2001. World Health Report 2000: responses to Murray and Frenk. Lancet 357 (9269), 1701–1702.
Nepomnyaschy, L., Reichman, N.E., 2006. Low birthweight and asthma among young urban children. Am. J. Public Health 96, 1604–1610.
Nilsson, P., 2009. The Long-Term Effects of Early Childhood Lead Exposure: Evidence from the Phase-Out of Leaded Gasoline. Uppsala University, Uppsala.
Nilsson, T., Bergh, A., 2013. Income inequality, health and development: in search of a pattern. In: Rosa Dias, P., O'Donnell, O. (Eds.), Health and Inequality. Emerald, Bingley, UK.
O'Donnell, O., 1995. Labour supply and saving decisions with uncertainty over sickness. J. Health Econ. 14 (4), 491–504.
Oecd, 2010. Sickness, Disability and Work: Breaking the Barriers: A Synthesis of Findings Across OECD Countries. OECD Publishing, Paris.
Oreopoulos, P., Stabile, M., Walld, R., Roos, L.L., 2008. Short-, medium-, and long-term consequences of poor infant health: an analysis using siblings and twins. J. Hum. Resour. 43 (1), 88–138.
Paneth, N., Susser, M., 1995. Early origin of coronary heart disease (the "Barker hypothesis"). Br. Med. J. 310 (6977), 411–412.
Parsons, D.O., 1980. The decline in male labor force participation. J. Polit. Econ. 88 (1), 117–134.
Pascual, M., Cantarero, D., Sarabia, J., 2005. Income inequality and health: do the equivalence scales matter? Atlantic Econ. J. 33 (2), 169–178.
Paxson, C., Schady, N., 2005. Child health and economic crisis in Peru. World Bank Econ. Rev. 19 (2), 203–223.
Paxson, C., Schady, N., 2010. Does money matter? The effects of cash transfers on child development in rural Ecuador. Econ. Dev. Cult. Change 59 (1), 187–229.
Pelkowski, J.M., Berger, M.C., 2004. The impact of health on employment, wages, and hours worked over the life cycle. Q. Rev. Econ. Finance 44 (1), 102–121.
Persico, N., Postlewaite, A., Silverman, D., 2004. The effect of adolescent experience on labor market outcomes: the case of height. J. Polit. Econ. 112 (5), 1019–1053.
Petrou, S., Kupek, E., 2008. Social capital and its relationship with measures of health status: evidence from the Health Survey for England 2003. Health Econ. 17 (1), 127–143.
Phelps, E.S., 1972. The statistical theory of racism and sexism. Am. Econ. Rev. 62, 659–661.
Pollitt, E., Gorman, K.S., Engle, P.L., Martorell, R., Rivera, J., 1993. Early Supplementary Feeding and Cognition: Effects Over Two Decades. University of Chicago Press, Chicago.
Power, C., Matthews, S., Manor, O., 1996. Inequalities in self rated health in the 1958 birth cohort: lifetime social circumstances or social mobility? Br. Med. J. 313 (7055), 449–453.
Power, C., Matthews, S., Manor, O., 1998. Inequality in self-rated health: explanations from different stages of life. Lancet 351, 1009–1014.
Preston, S.H., 1975. The changing relation between mortality and level of economic development. Popul. Stud. 29 (2), 231–248.
Pritchett, L., Summers, L.H., 1996. Wealthier is healthier. J. Hum. Resour. 31 (4), 841–868.
Propper, C., Rigg, J., Burgess, S., 2007. Child health: evidence on the roles of family income and maternal mental health from a UK birth cohort. Health Econ. 16 (11), 1245–1269.
Putnam, R.D., Leonardi, R., Nanetti, R., 1993. Making Democracy Work: Civic Traditions in Modern Italy. Princeton University Press, Princeton, NJ.
Regidor, E., Martínez, D., Santos, J.M., Calle, M.E., Ortega, P., Astasio, P., 2012. New findings do not support the neomaterialist theory of the relation between income inequality and infant mortality. Soc. Sci. Med. 75 (4), 752–753.
Reinhold, S., Jürges, H., 2012. Parental income and child health in Germany. Health Econ. 21 (5), 562–579.
Relton, C.L., Davey Smith, G., 2012. Is epidemiology ready for epigenetics? Int. J. Epidemiol. 41 (1), 5–9.

Riley, J.C., 1989. Sickness, Recovery and Death: A History and Forecast of Ill-Health. Macmillan, Basingstoke.
Riphan, R., 1999. Income and employment effects of health shocks. A test case for the German welfare state. J. Popul. Econ. 12 (3), 363–389.
Rivera, J.A., Sotres-Alvarez, D., Habicht, J., Shamah, T., Villalpando, S., 2004. Impact of the Mexican program for education, health, and nutrition (Progresa) on rates of growth and anemia in infants and young children. JAMA 291 (21), 2563–2570.
Rodgers, G.B., 1979. Income and inequality as determinants of mortality: an international cross-section analysis. Popul. Stud. 33, 343–351.
Ronconi, L., Brown, T.T., Scheffler, R.M., 2012. Social capital and self-rated health in Argentina. Health Econ. 21 (2), 201–208.
Royer, H., 2009. Separated at girth: estimating the long-run and intergenerational effects of birthweight using twins. Am. Econ. J. 1 (1), 49–85.
Ruhm, C.J., 2000. Are recessions good for your health? Q. J. Econ. 115 (2), 617–650.
Ruhm, C.J., 2003. Good times make you sick. J. Health Econ. 22 (4), 637–658.
Ruhm, C.J., 2013. Recessions, healthy no more? NBER Working Paper 19287, Cambridge, MA.
Saigal, S., Szatmari, P., Rosenbaum, P., Campbell, D., King, S., 1991. Cognitive abilities and school performance of extremely low birth weight children and matched term control children at age 8 years: a regional study. J. Pediatr. 118 (5), 751–760.
Salm, M., 2011. The effect of pensions on longevity: evidence from Union Army Veterans. Econ. J. 121 (552), 595–619.
Sapolsky, R.M., 2005. The influence of social hierarchy on primate health. Science 308 (5722), 648–652.
Sastre, M., Trannoy, A., 2002. Shapley inequality decomposition by factor components: some methodological issues. J. Econ. 9 (1), 51–89.
Saunders, P.R., Evans, N., 2010. Beware False Prophets: Equality, the Good Society and the Spirit Level. Policy Exchange, London.
Schmeiser, M.D., 2009. Expanding wallets and waistlines: the impact of family income on the BMI of women and men eligible for the Earned Income Tax Credit. Health Econ. 18 (11), 1277–1294.
Schultz, T.P., 2005. Productive Benefits of Health: Evidence from Low-Income Countries. Yale University, New Haven.
Schultz, T.P., 2010. Health human capital and economic development. J. Afr. Econ. 19 (Suppl. 3), iii12–iii80.
Schwandt, H., 2013. Wealth Shocks and Health Outcomes. Center for Health and Wellbeing, Princeton University, Mimeo. http://www.princeton.edu/∼schwandt/.
Semykina, A., Wooldridge, J.M., 2010. Estimating panel data models in the presence of endogeneity and selection. J. Econometrics 157 (2), 375–380.
Sen, A., 1999. Development as Freedom. Knopf, New York.
Sheps, M.C., 1961. Marriage and mortality. Am. J. Public Health 51, 547–555.
Shorrocks, A., 2013. Decomposition procedures for distributional analysis: a unified framework based on the Shapley value. J. Econ. Inequal. 11 (1), 99–126.
Sickles, R.C., Taubman, P., 1986. An analysis of the health and retirement status of the elderly. Econometrica 54 (6), 1339–1356.
Sickles, R.C., Yazbeck, A., 1998. On the dynamics of demand for leisure and the production of health. J. Bus. Econ. Stat. 16 (2), 187–197.
Siegel, M., 2006. Measuring the effect of husband's health on wife's labour supply. Health Econ. 15, 579–601.
Sloan, F.A., Hsieh, C-R., 2012. Health Economics. MIT Press, Cambridge, MA.
Smith, J., 1999. Healthy bodies and thick wallets: the dual relation between health and socioeconomic status. J. Econ. Perspect. 13, 145–166.
Smith, J.P., 2004. Unraveling the SES: health connection. Popul. Dev. Rev. 30, 108–132.
Smith, J.P., 2007. The impact of socioeconomic status on health over the life-course. J. Hum. Resour. XLII (4), 739–764.

Smith, J.P., 2009. The impact of childhood health on adult labor market outcomes. Rev. Econ. Stat. 91 (3), 478–489.
Snowdon, C., 2010. The Spirit Level Delusion: Fact-Checking the Left's New Theory of Everything. Little Dice, Ripon.
Snyder, S.E., Evans, W.N., 2006. The effect of income on mortality: evidence from the social security notch. Rev. Econ. Stat. 88 (3), 482–495.
Steckel, R.H., 1995. Stature and the standard of living. J. Econ. Lit. 33 (4), 1903–1940.
Stern, S., 1989. Measuring the effect of disability on labor force participation. J. Hum. Resour. 24, 361–395.
Stevens, A.H., Miller, D.L., Page, M.E., Filipski, M., 2011. The Best of Times, the Worst of Times: Understanding Pro-Cyclical Mortality. NBER Working Paper No. 17657, National Bureau of Economic Research, Cambridge, MA.
Stiglitz, J.E., 1976. The efficiency wage hypothesis, surplus labour, and the distribution of income in L.D.C.s. Oxf. Econ. Papers 28 (2), 185–207.
Stowasser, T., Heiss, F., Mcfadden, D.L., Winter, J., 2012. "Healthy, Wealthy and Wise?" Revisited: an analysis of the causal pathways from socio-economic status to health. In: Wise, D.A. (Ed.), Investigations in the Economics of Aging. University of Chicago Press, Chicago, pp. 267–317.
Strauss, J., Thomas, D., 1998. Health, nutrition, and economic development. J. Econ. Lit. 36 (2), 766–817.
Subramanian, S.V., Kawachi, I., 2004. Income inequality and health: what have we learned so far? Epidemiol. Rev. 26 (1), 78–91.
Tacke, T., Waldmann, R.J., 2013. Infant mortality, relative income and public policy. Appl. Econ. 45 (22), 3240–3254.
Tekin, E., McClellan, C., Minyard, K.J., 2013. Health and Health Behaviors During the Worst of Times: Evidence from the Great Recession. NBER Working Paper No. 19234, Cambridge, MA.
Theodossiou, I., Zangelidis, A., 2009. The social gradient in health: the effect of absolute income and subjective social status assessment on the individual's health in Europe. Econ. Hum. Biol. 7 (2), 229–237.
Thomas, D., Frenkenberg, E., 2002. Health, nutrition and prosperity: a microeconomic perspective. Bull. World Health Organ. 80 (2), 106–113.
Thorbecke, E., Charumilind, C., 2002. Economic inequality and its socioeconomic impact. World Dev. 30 (9), 1477–1495.
Townsend, R.M., 1994. Risk and insurance in village India. Econometrica 62 (3), 539–591.
Twaddle, A., 1974. The concept of health status. Soc. Sci. Med. 8, 29–38.
UNESCO Institute for Statistics, ISCED, International Standard Classification of Education. Available http://www.uis.unesco.org.
Üstün, T., Chatterji, S., Villanueva, M., Bendib, L., Çelik, C., Sadana, R., Valentine, N., Ortiz, J., Tandon, A., Salomon, J., Cao, Y., Xie, W., Özaltin, E., Mathers, C., Murray, C., 2003. WHO multi-country survey study on health and responsiveness, 2000–2001. In: Murray, C.J.L., Evans, D.B. (Eds.), Health Systems Performance Assessment: Debates, Methods and Empiricism. World Health Organization, Geneva, pp. 761–796.
Van Doorslaer, E., Van Ourti, T., 2011. Measuring inequality and inequity in health and health care. In: Glied, S., Smith, P.C. (Eds.), Oxford Handbook of Health Economics. Oxford University Press, Oxford, pp. 837–869.
Van Doorslaer, E., Wagstaff, A., Bleichrodt, H., Calonge, S., Gerdtham, U.G., Gerfin, M., Geurts, J., Gross, L., Hakkinen, U., Leu, R.E., O'Donnell, O., Propper, C., Puffer, F., Rodriguez, M., Sundberg, G., Winkelhake, O., 1997. Income-related inequalities in health: some international comparisons. J. Health Econ. 16 (1), 93–112.
Van Doorslaer, E., Wagstaff, A., Van Der Burg, H., Christiansen, T., De Graeve, D., Duchesne, I., Gerdtham, U., Gerfin, M., Geurts, J., Gross, L., 2000. Equity in the delivery of health care in Europe and the US. J. Health Econ. 19 (5), 553–583.
Van Doorslaer, E., Koolman, X., Jones, A.M., 2004. Explaining income-related inequalities in doctor utilization in Europe. Health Econ. 13 (7), 629–647.
Van Doorslaer, E., Masseria, C., Koolman, X., 2006. Inequalities in access to medical care by income in developed countries. Can. Med. Assoc. J. 174 (2), 177–183.

Van Groezen, B., Jadoenandansing, R., Pasini, G., 2011. Social capital and health across European countries. Appl. Econ. Lett. 18 (12), 1167–1170.
Van Houtven, C., Coe, N., 2010. Spousal Health Shocks and the Timing of the Retirement Decision in the Face of Forward-Looking Financial Incentives. Center for Retirement Research at Boston College, Boston, Working Paper WP 2010-7.
Van Kippersluis, H., Galama, T.J., 2013. Why the Rich Drink More But Smoke Less: The Impact of Wealth on Health Behaviors. DP 13-035/V, Tinbergen Institute, Amsterdam/Rotterdam.
Wadsworth, M., Kuh, D., 1997. Childhood influences on adult health: a review of recent work for the British 1946 Birth Cohort Study, the MRC National Survey of Health and Development. Paediatr. Perinat. Epidemiol. 11, 2–20.
Wagstaff, A., Van Doorslaer, E., 2000. income inequality and health: what does the literature tell us? Annu. Rev. Public Health 21 (1), 543–567.
Wagstaff, A., Paci, P., Vandoorslaer, E., 1991. On the measurement of inequalities in health. Soc. Sci. Med. 33 (5), 545–557.
Waldmann, R.J., 1992. Income distribution and infant mortality. Q. J. Econ. 107, 1283–1302.
Waldron, I., Hughes, M.E., Brooks, T.L., 1996. Marriage protection and marriage selection—prospective evidence for reciprocal effects of marital status and health. Soc. Sci. Med. 43 (1), 113–123.
Walker, S.P., Chang, S.M., Powell, C.A., Grantham-Mcgregor, S.M., 2005. Effects of early childhood psychosocial stimulation and nutritional supplementation on cognition and education in growth-stunted Jamaican children: prospective cohort study. Lancet 366, 1804–1807.
Weaver, I.C.G., Cervoni, N., Champagne, F.A., D'Alessio, A.C., Sharma, S., Seckl, J.R., Dymov, S., Szyf, M., Meaney, M.J., 2004. Epigenetic programming by maternal behavior. Nat. Neurosci. 7 (8), 847–854.
West, P., 1997. Unhealthy societies: the afflictions of inequality (book). Sociol. Health Illn. 19 (5), 668–670.
Wildman, J., Gravelle, H., Sutton, M., 2003. Health and income inequality: attempting to avoid the aggregation problem. Appl. Econ. 35 (9), 999–1004.
Wilkinson, R.G., 1990. Income distribution and mortality: a 'natural' experiment. Sociol. Health Illn. 12 (4), 391–412.
Wilkinson, R.G., 1992. Income distribution and life expectancy. Br. Med. J. 304, 165–168.
Wilkinson, R.G., 1996. Unhealthy Societies: the Afflictions of Inequality. Routledge, London.
Wilkinson, R.G., 2001. Mind the Gap: Hierarchies, Health and Human Evolution. Yale University Press, New Haven.
Wilkinson, R.G., Pickett, K.E., 2006. Income inequality and population health: a review and explanation of the evidence. Soc. Sci. Med. 62 (7), 1768–1784.
Wilkinson, R.G., Pickett, K., 2010. The Spirit Level: Why Equality Is Better for Everyone. Penguin Books, London/New York.
Wise, D.A. (Ed.), 2012. Social Security Programs and Retirement Around the World: Historical Trends in Mortality and Health, Employment, and Disability Insurance Participation and Reforms. The University of Chicago Press, Chicago.
Woodbury, R.M., 1924. Economic factors in infant mortality. J. Am. Stat. Assoc. 19, 137–155.
Wu, S., 2003. The effects of health events on the economics status of married couples. J. Hum. Resour. 38 (1), 219–230.
Yang, T., Chen, V.Y., Shoff, C., Matthews, S.A., 2012. Using quantile regression to examine the effects of inequality across the mortality distribution in the U.S. counties. Soc. Sci. Med. 74 (12), 1900–1910.
Yitzhaki, S., 1979. Relative deprivation and the Gini coefficient. Q. J. Econ. 93 (2), 321–324.
Yitzhaki, S., 1983. On an extension of the Gini inequality index. Int. Econ. Rev. 24 (3), 617–628.

CHAPTER 18

Labor Market Institutions and the Dispersion of Wage Earnings

Wiemer Salverda*,†, Daniele Checchi‡
*Amsterdam Institute for Advanced Labour Studies (AIAS), University of Amsterdam, Amsterdam, The Netherlands
†Amsterdam Centre for Inequality Studies (AMCIS), University of Amsterdam, Amsterdam, The Netherlands
‡Department of Economics, University of Milan, Milano, Italy and IZA, Bonn, Germany

Contents

Handbook of Income Distribution, Volume 2B
ISSN 1574-0056, http://dx.doi.org/10.1016/B978-0-444-59429-7.00019-4

Abstract

Considering the contribution of the distribution of individual wages and earnings to that of household incomes we find two separate literatures that should be brought together, and bring “new institutions” into play. Growing female employment, rising dual-earnership and part-time employment underline its relevance. We discuss the measurement of wage inequality, data sources, and stylized facts of wage dispersion for rich countries. The literature explaining the dispersion of wage rates and the role of institutions is evaluated, from the early 1980s to the recent literature on job polarization and tasks as well as on the minimum wage. Distinguishing between supply-and-demand approaches and institutional ones, we find supply and demand challenged by the empirical measurement of technological change and a risk of ad hoc additions, without realizing their institutional preconditions. The institutional approach faces an abundance of institutions without a clear conceptual delineation of institutions and their interactions. Empirical cross-country analysis of the correlation between institutional measures and wage inequality incorporates unemployment and working hours dynamics, discussing the problems of matching individuals to their relevant institutional framework. Minimum wage legislation and active labor market policies come out negatively correlated to earnings inequality in US and EU countries.

Keywords

Labor market institutions, Household labor supply, Hourly wages, Hours worked, Annual earnings, Dispersion, Inequality measures, Household incomes, Minimum wage, Unions, Employment protection

JEL Classification Codes

D02, D13, D31, J22, J31, J51, J52

18.1. INTRODUCTION

This is not “simply” a study of the literature regarding wage inequality in the labor market, even apart from the fact that the literature is immense. The income distribution is the focus of the present handbook and provides the ultimate rationale for considering the dispersion of wage earnings here. It is natural therefore to consider the distribution of individual wages and earnings in the labor market in light of what it may contribute to the distribution of household incomes, which are the common unit of analysis for the income distribution. One may surmise that the subject of how wage inequality and income inequality relate has gained relevance—and also complexity—as the growing labor market participation of women and the concomitant rise of dual-earner households make societies move away from the single-earner breadwinner model, in which labor market earnings closely resemble household income.[1] The recent literature on household joblessness provides further encouragement. Nevertheless, the two strands of study, of wage dispersion on the one hand and household income distribution on the other, are miles apart. There is a growing literature aiming to measure the distance between the two distributions and attempting to bring them together, but it is still small and also rather diverse. More importantly, there is very little in this literature that also accounts for the role of institutions with respect to the interrelationship between the two distributions, though that role will be significant as one can infer from the burgeoning literature on institutions and female labor supply. In addition, these are often new institutions (e.g., parental leave, child care arrangements, job entitlements during maternity leave, and/or changing from full-time to part-time employment), which seem deserving of attention together with the traditional labor market institutions (LMIs) (minimum wage, employment protection, union density, etc.).

However, understanding institutions in relation to wage dispersion is our overarching purpose—and a very demanding purpose in its own right. It would require a bridge too far to also try to overcome the gap and incorporate the income distribution in our approach. Instead we will take a swift look at said literature and the stylized facts of the subject, and we will do so at the start of our argument to make the best of it as a heuristic device for our ensuing discussion of wage dispersion and institutions. Thus, we hope to make a contribution on which future analysis can expand by providing a building block that can be used subsequently for constructing a unified economic theory of income distribution, a theory that is still missing (Atkinson and Bourguignon, 2000, 26). By the way, that building block itself needs to account for the very fact that neither a unified theory of earnings dispersion is available. We intend to do that by reviewing the literature on institutions and earnings distribution in a framework that may be relevant also for further use in studying the household income distribution.

Concretely, we explicitly include in our focus the distribution of *annual* earnings from labor as the income distribution is commonly measured and analyzed on an annual basis.[2] This entails, first, that we study both wage rates and (annual) hours of work—which

[1] In a world of joint within-household labor supply, the two distributions will deviate from each other unless households supply the same number of hours, and wage rates are identical across household members.

[2] We will be more precise about such concepts in Section 18.3.

taken together make up annual earnings—as well as the dispersion of both and their interrelationship. Thus, we aim to go beyond, e.g., Blau and Kahn (1999), who address the effects of wage-setting institutions on wage inequality as well as on employment, but for the latter restrict themselves to aggregate employment effects and ignore its dispersion over individuals and households as well as its relationship to wage dispersion. It implies that one needs to consider the role that institutions play not only in relation to the wage rate, the hours worked, and the individual probability of employment, but ultimately also in relation to the household distribution of employment—what we can call a double-edged employment perspective. At the same time, this brings into play the role of unemployment and joblessness (zero hours), the frequency of which may also be affected by institutions. More generally, individual institutions that primarily concern one of these aspects, say the wage rate, will need to be considered also in relation to the other aspects. The separate effects may differ and in the end it is their joint effect that counts.[3]

Second, we will contemplate the relevance and the effects of LMIs from this distribution point of view. Particularly, we will on the one hand leave aside the literature that focuses on wage dispersion in relation to the matching of workers to given jobs (e.g., Mortensen, 2005, on search, or Rosen, 1986, on compensating differentials). We also leave out the literature on other important facets of inequality such as earnings mobility or its role as a work and career incentive. On the other hand we will look—to the extent that we can—for institutions that may affect the distribution of employment over households (e.g., equal treatment, working-hours nondiscrimination, child care provisions or tax measures) or the supply of hours over the year (e.g., temp agency work, temporary contracts). Thus, different institutions from the usual suspects may come into play, for example, new rules and regulations regarding part-time jobs and pay, or the "reconciliation of work and family life," while at the same time those usual suspects will be checked for their effects in this domain. The "new" institutions will need to be considered in their own right but, naturally, also in relation to the previous ones. We need to be careful, though, that the assortment of institutions under scrutiny be manageable; as in modern society labor market behavior has become so central to human existence that virtually any institution might be thought to have an effect.

In our take on the literature, we aim to be careful in considering the role of institutions not in isolation of the "normal" economy. That is, we may compare, for example, the meticulous evaluation of the literature by Katz and Autor (1999), who first discuss the role of supply and demand and after that turn to institutions, or the warning given by Blau and Kahn (1999, 1416) with regard to international comparative studies of the effects of institutions "that many things besides the institutions in question may differ across countries, so we cannot be certain if the institutions are really responsible for the observed differences in outcomes." Similarly, we need to remain aware of noninstitutional effects

[3] Interestingly, Olivetti and Petrongolo (2008) consider the interaction of hours and wages from a different perspective, focusing on the effect on the gender pay gap of the number of women being employed.

influencing market labor supply, such as, e.g., technical progress in household production (cf. Kahn, 2005). More generally, we sympathize with Manning (2011), who prefers to phrase his recent overview not in terms of canonical models, where "precision relates to the models and not the world and can easily become spurious precision when the models are very abstract with assumptions designed more for analytical tractability than realism" (2011, 975). In our view, the distribution of earnings is very much a phenomenon of crucial importance in "the world." Though we aim to broaden the scope to include the dispersion of employment, we do not and cannot possibly pursue this in a general equilibrium format. Further to this, being aware of significant differences among countries, we leave open the possibility that one size may not fit all.

In addition, we like to stress that the time period effectively covered in the chapter is determined by the literature that we aim to address. Though that period may seem long to some as we begin our coverage at the end of the 1960s for certain countries, it is important to realize that the trends found may be selective. The long-run historical perspectives adopted in the top-incomes literature (Alvaredo et al., 2013) or in Atkinson's (2008) internationally comparative study of the earnings distribution suggest that preceding trends may diverge, sometimes radically, and might throw a different light on the mechanisms at work. Ultimately, this may tell a different story, but the study of this is in its infancy.

Before continuing, we mention a caveat regarding the two concepts of "dispersion" and "inequality," which we have used indiscriminately to indicate the squeeze or stretch of a distribution. A major reason for many to pay attention to the dispersion is that a large part of it coincides with social or economic inequality as it is commonly understood. However, more precisely, the dispersion is thought to relate to a range of observations, wages, or incomes in this case, that are not all the same and therefore are unequal in a mechanical, mathematical sense of the word. Inequality, by contrast, provides a qualifier to such observations that makes them unequal in the sense of analyses providing an explanatory interpretation of the observations, either individual or aggregated. So, strictly speaking, dispersion and inequality are different concepts. Not all mechanical differences will also be inequalities from an analytical point of view, for example, differences in individual earnings that reflect differences in efforts. Conversely, not all analytical inequalities will also be mechanical differences, for example, individual earnings that are identical in spite of differences in efforts. Having said this we will continue to use the two words interchangeably as this chapter is aimed at evaluating a set of such qualifying analyses. Note, finally, that measures of dispersion or of inequality (Gini coefficient, etc.) are identical, and are usually called measures of inequality—terms that we will also use in this chapter.

Some of the above references indicate the existence of various literature overviews that are relevant to our study of earnings inequality, which are found in the first volume of the *Handbook of Income Distribution*, all volumes of the *Handbook of Labor Economics*, and the *Oxford Handbook of Economic Inequality*. We will not redo these, but gratefully build on them when it is useful to do so. Note that not only economists but also political and social scientists have studied the subject (Alderson and Nielsen, 2002; Becher and Pontusson, 2011; DiPrete, 2007; Golden and Wallerstein, 2011; Kenworthy and Pontusson, 2005; Oliver, 2008; Wallerstein, 1999). We will also allude to some of their results.

Our contribution takes the general level of inequality as its starting point but cannot escape digging below that surface. Thus, for example, we may touch upon the tails of the distribution—top incomes, (in-work) poverty—where much of the action is. However, for a deeper understanding of those tails as well as the complementing middle we refer to the treatment of polarization (Chapter 5), top incomes (Chapter 7), and in-work poverty (Chapter 23) elsewhere in this handbook. More generally, the labor market also figures as one of the multiple causes of inequality in Chapter 19. On another dimension, our contribution stops short of the within-household distribution (see Chapter 16) or any further analysis of gender inequality (see Chapter 12). Finally, this chapter will cover those countries that have well-developed, comprehensive formal labor markets. This restricts the selection of the literature to analyses that concern the United States, Canada, Japan, Korea, Australia, New Zealand, the member states of the European Union, and some other European countries such as Iceland, Norway, and Switzerland.

18.1.1 Lay Out

The layout of the chapter is as follows. First, in Section 18.2, we will briefly discuss the literature that regards the link between wage dispersion and the household income distribution, considering the distributions of earnings and employment from both the individual and the household perspectives, and presenting some stylized facts. In Section 18.3 we discuss the measurement of wage inequality with some relevant data sources and present some stylized facts of wage dispersion for a selection of countries. Next, in Section 18.4, we discuss theories aimed at explaining the dispersion of wage rates and the role of institutions. Section 18.5 then addresses the role of LMIs empirically, with the help of a model that that incorporates several features advocated in the preceding sections, such as a focus on earnings, i.e., the product of wage rate and annual efforts, and that inserts as explanatory variables a number of "new institutions" related to household labor supply. In addition, we use recent internationally comparative data. Finally, we conclude in Section 18.6 by summarizing the main findings and considering issues warranting further research.

18.2. EARNINGS DISTRIBUTION AND INCOME DISTRIBUTION: A SHORT TALE OF TWO LONG LITERATURES

In spite of recent declines in the labor share in GDP or national income,[4] the income that people generate in the labor market is obviously the most frequent and most important

[4] We leave aside here the relationship between the labor share in GDP (declining in many countries) and the income distribution. Compare, e.g., Atkinson (2009), Glyn (2009), Checchi and Garcia-Peñalosa (2008), and OECD (2012).

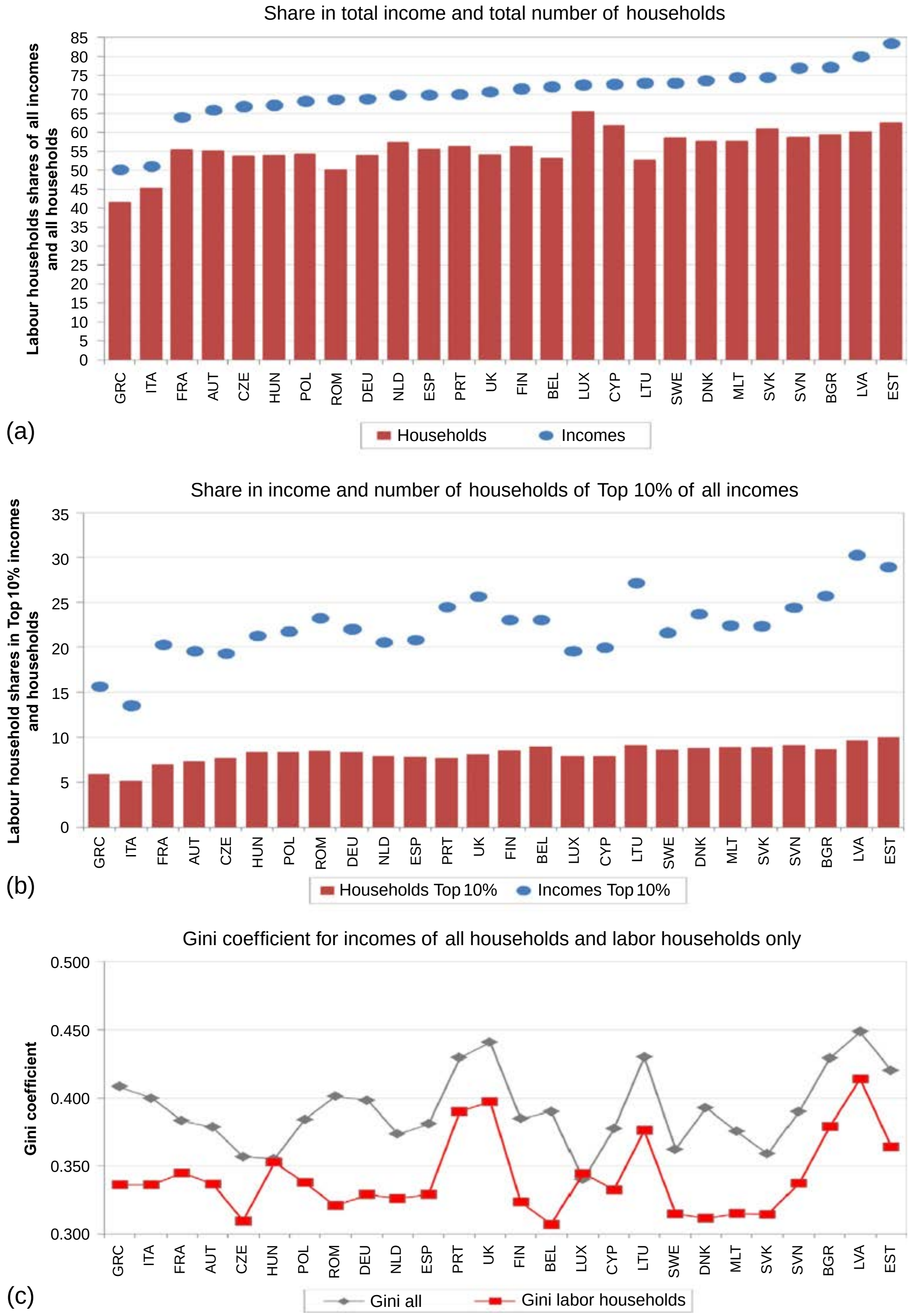


Figure 18.1 See legend on next page.

part of household incomes, and the inequality of labor earnings seems an important determinant of income inequality at face value. Figure 18.1 portrays in three panels the role of "labor households," which are defined as households receiving more than half their total income from wage earnings, across 26 countries of the European Union. Panel (a) ranks the countries by the income share of labor households (the markers), and the same ranking is adopted for the other panels. Panel (a) indicates that labor households receive the majority of all incomes, ranging from slightly over 50% in Greece and Italy up to a maximum of 84% in Estonia. They comprise significantly smaller shares of all households, however, ranging from less than half in Greece and Italy[5] to 66% in Luxembourg. Clearly, these households' mean incomes are above average in all countries. This is borne out by panel (b), which indicates similar shares with a focus on the Top 10% of all incomes in a country. The income share always exceeds the household share and does so by far: on average the income share is 14% points higher than the household share. This contrast with the Bottom 90% (not shown): here the gaps between the two shares are modest, and they can be positive as well as negative; the resulting cross-country average is almost nil. At the same time, in panel (c), the Gini coefficients for all households always exceed that for labor households and they move in striking parallel in various countries characterized by high labor-income inequality such as the United Kingdom, Portugal, Lithuania, Bulgaria, Latvia, and Estonia (overall correlation is 0.75). The Gini levels do not follow the smooth ranking of increasing income shares but vary substantially (correlation 0.23). Therefore, rather dissimilar Gini coefficients can go together with very similar income shares as the middle group, ranging from Germany to Belgium, as illustrated (panel c vs. a). However, for labor households income shares in the Top 10% and the Gini coefficient show a more similar pattern (panel c compared to panel b) (correlation 0.56). So income from labor is highly important indeed, but its effects on income inequality show significant variation and warrant further scrutiny.

Figure 18.1 Importance of labor households and their annual incomes, 26 European countries ranked by total income share, 2010. (a) Share in total income and total number of households. (b) Share in income and number of households of Top 10% of all incomes. (c) Gini coefficient for incomes of all households and labor households only. *Reading note*: In Greece labor households receive 50% of all incomes and make up 42% of all households; among them 6% have an income in the Top 10% of the overall income distribution receiving 16% of all incomes; the Greek Gini coefficient for all incomes is 0.408 while for labor households it is 0.336. *Explanatory note*: Labor households derive more than 50% of their total income from wage earnings. We use the ISO 3-alpha country codes in all relevant graphs (see list in Appendix A). Unfortunately, data for Ireland are not available. Source*: Calculated from EU-SILC 2011 (compare Salverda and Haas (2014) for a comparison for the working-age bracket only).*

[5] Note that the low household share largely explains the low income share.

Another measure of inequality, the income share of the top decile of the distribution, tells basically the same story for all incomes as the Gini and the top share are highly correlated (0.91) (compare Leigh, 2009). However, the gap between all incomes and labor incomes is more substantial here—the correlation of the two top shares is only 0.32—and suggests that the role of high levels of household earnings differs significantly between countries. The linkage between the dispersion of wages and the income distribution is clearly important and also warrants further research.

Though the literature on the two distributions is not absent and perhaps even growing, it is not the subject of a strong strand. Instead, one may surmise, there are two largely separate, extensive literatures, one addressing (individual) wage inequality in the labor market and the other (household) income inequality in society. As Gottschalk and Danziger (2005, 253) observe "Labor economists have tended to focus on changes in the distribution of wage rates, the most restrictive income concept, since they are interested in changes in market and institutional forces that have altered the prices paid to labor of different types. At the other extreme, policy analysts have focused on changes in the distribution of the broadest income concept, family income adjusted for family size. This reflects their interest in changes in resources available to different groups, including the poor." It confirms that the conclusion drawn 8 years before by Gottschalk and Smeeding (1997, 676), that "an overall framework would simultaneously model the generation of all sources of income ... as well as the formation of income sharing units" and be considered "the next big step that must be taken," was still a tall order when Gottschalk and Danziger made their contribution. Yet another 5 years later, Jiří Večerník (2010, 2) observed that "there seems to be a gulf between the analysis of personal earnings and household income." It seems a foregone conclusion that for the combination of individual wage and earnings inequality and household earnings and income inequality, the unified economic theory of income distribution, hoped for by Atkinson and Bourguignon (2000, 26), is not yet forthcoming though interesting contributions may be found below.[6]

This divide has a technical aspect that deserves some attention. The dispersion of wages is commonly conceived as the distribution of hourly wages, i.e., wage rates. The income distribution, by contrast, focuses on annual incomes, and therewith annual earnings, which are the product of hourly wages and annual hours worked. Next to the wage distribution, this brings into play the distribution of hours worked during the year, which, in turn, are the product of jobs and hours on the job. These hours have become a significant dimension of employment in many countries because of the growing importance of part-time employment and temporary jobs. Their presence adds to the traditional effect on annual hours that is exerted by the turnover during the year of people who join or leave

[6] However, for a number of developing countries (which are not the subject of this chapter) a valuable attempt with interesting results has been made by Bourguignon et al. (2004) in decomposing household income inequality changes along the relevant dimension of labor market behavior and outcomes.

employment.[7] As a result we deem it essential to distinguish between various distributions: *wages* (which are hourly), *earnings* (which are annual), *employment* (which concerns annual hours worked), and *incomes* (which include other sources than earnings).

A second difference is that the wage distribution is commonly conceived in *gross* terms, that is pretax, whereas on the income side there is a strong focus on *disposable* incomes—after transfers and taxes—which are often also standardized (equivalized) for the size and composition of the receiving household.[8] The third difference is that the dispersion of wages rests on the individual as the unit of analysis whereas the income distribution is based on the household, which can be a combination of individuals. Thus, for linking the two distributions, the individuals from the one side need to be linked to their households on the other side. Importantly, this puts the limelight on the distribution of employment and corresponding earnings over households. There is a significant literature on the other side of this employment coin, the nonemployment or joblessness of households, especially in comparison to individual joblessness, which was started by Paul Gregg and Jonathan Wadsworth in the mid-1990s (Gregg et al., 1996, 1998 and Gregg and Wadsworth, 2008). However, this literature is not often linked to the distribution of incomes albeit it may be linked to poverty (De Graaf-Zijl and Nolan, 2011).

18.2.1 Individual or Household Incomes?

Before discussing the main points found in the literature we present a few stylized facts that may demonstrate the relevance of considering the link between the two distributions. First, we consider the employment side of the matter. A core message from the joblessness literature is that in many countries individual workless rates have fallen over the past 20 years, but household-based workless rates have not (Gregg et al., 2010, 161). Or to put it the other way around, the growth in (individual) employment-to-population ratios has not been mirrored in a corresponding increase in what can be termed the "household employment rate." The implication is that much of the additional jobs growth has gone to households already containing a worker. Figure 18.2 illustrates this for a number of European countries since the mid-1990s: most of the decline in individual unemployment has gone to households already engaged in employment and much less has contributed to a lowering of the number of people living in jobless households.

[7] Including temporary employment of less than 1 year this is reflected in the difference between the distributions of the full-year and the part-year employed. For example, Salverda et al. (2013), in Figure 2.11, shows for the Netherlands that the P90:P10 percentile ratio is halved when attention shifts from all earners to full-year earners only.

[8] Equivalization serves to account for the demands that household members put on income as well as the economies of scale of jointly managing a household (Atkinson et al., 1995; Förster, 1994; OECD, 2009). Note that applying equivalization not only to disposable incomes but also to market incomes and gross incomes (e.g., OECD, 2011; and various contributions to the special issue of the Review of Economic Dynamics), may affect the perception of labor market outcomes on the one hand and changes between these three distributions on the other hand.

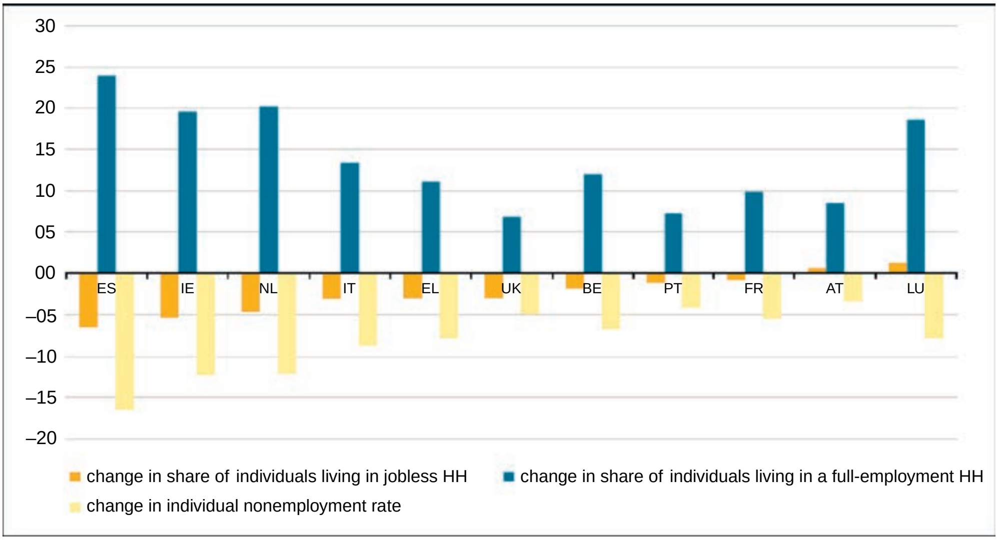


Figure 18.2 Changes (percentage points) in individual and household employment, 11 European countries, 1995–2008. *Reading note*: In Spain the share among individuals of those in work who are also members of a household where everyone is in work increased by 24% points between 1995 and 2008; the share for those living in households without work declined by 7% points; the share of individuals without work declined by 16.5% points. *Explanatory note*: In full-employment households everyone is in work; this includes single-person households. Employment follows the ILO definition and includes the self-employed. Persons aged 18–24 whose status is "inactive" are considered to be full-time students and excluded. For country codes see Appendix A. Source: *Eurostat—Corluy and Vandenbroucke, 2013, Figure 1 (based on the European Labour Force Survey).*

Figure 18.3 adds a particularly sharp example of the divergence between the two rates of employment for prime-age adults for the United Kingdom, one for persons (the traditional individual employment-to-population ratio), the other for households (the percentage of relevant households that have at least one employed person among their members). The former rate always exceeds the latter, and the gap between the two has grown rapidly from 2% points at the end of the 1970s to 13% points since the early 1990s.[9] Often such developments have gone hand in hand with an expansion of part-time employment. The correlation of individuals' levels of pay to their numbers of hours worked can tell us whether this hours-of-work dimension enhances or mitigates inequality. A positive correlation implies a more unequal distribution of annual earnings than of hourly earnings among individuals. The correlation has tended upward significantly and turned from negative to

[9] Atkinson (1993, 335 ff) discusses an 11.5% point decline in the family (adult) employment rate for the United Kingdom between 1975 and 1985 and infers that half the increase in inequality can be attributed to this "shift in work."

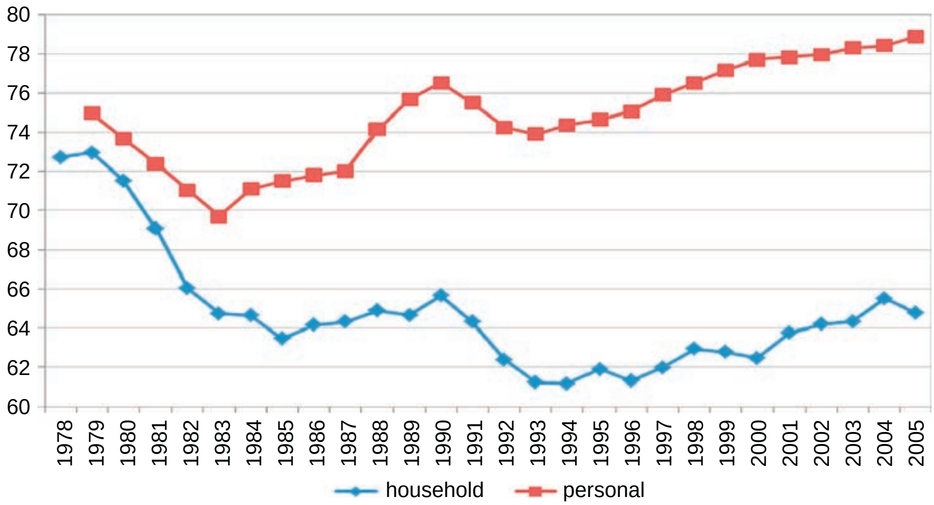


Figure 18.3 Employment rates (%) for individuals aged 25–59 and their households, United Kingdom, 1978–2005. *Reading note*: The share among individuals aged 25–59 who are in work grew from 75% in 1979 to 79% in 2005; the share of households corresponding to these individuals where at least one person is employed, declined from 73% to 65%. Source*: Derived from Blundell and Etheridge (2010), Figures 2.1 and 2.3 (based on Labor Force Survey and Family Expenditure Survey).*

positive in some countries although it still is negative in other countries. The correlation seems particularly strong for British women (Figure 18.4).

Compared to single-breadwinner households this complicates the relationship between the wage distribution and the income distribution. At the same time it makes the scrutiny of that relationship all the more important. Thus the role of dual-earner and multiple-earner households has expanded and is now substantial in many European countries as is indicated in Figure 18.5. With the exception of Italy and Greece, dual-earner and multiple-earner households are the majority among households, and evidently, employees in those households make up an even larger share of all employees. In particular, the role of multiple-earner households varies substantially across countries, from 4% of all households in Greece to 27% in Bulgaria.

In a world of full-time-working single-earner households, the correspondence between wage dispersion and income distribution seems pretty straightforward: a high individual wage directly implies a high household income. This traditional situation may provide another explanation, for lack of a problem, why the literature on the linkage between the two distributions seems underdeveloped. The formation of households and their labor supply may affect the distribution of incomes depending on the correlation of earnings levels between the earners in a household. A positive correlation will enhance household earnings inequality, in addition to the frequency of the occurrence of joint earnings. Changes in mating behavior or in partners' employment participation or both at the same time will be behind this. Figure 18.6 indicates the rise in the correlation

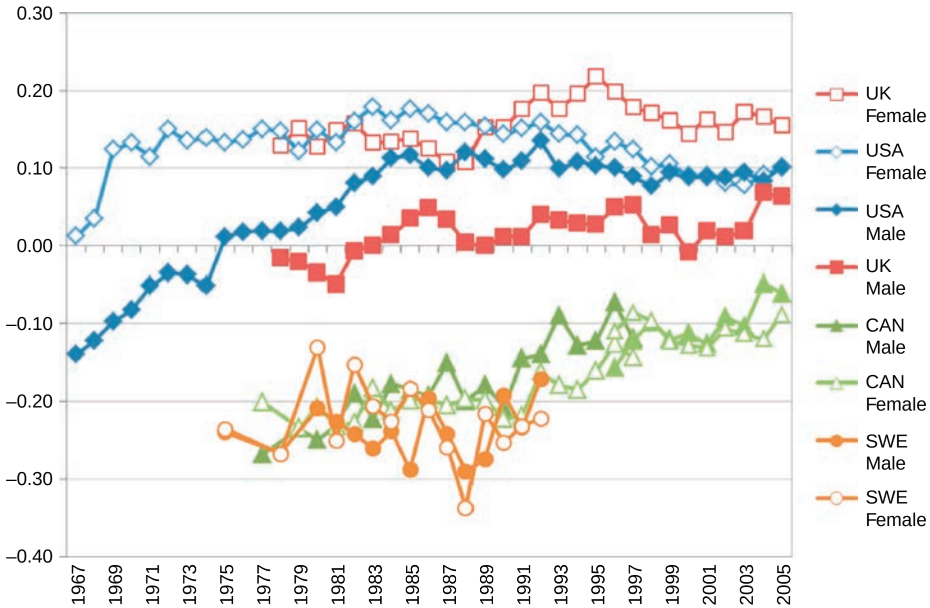


Figure 18.4 Correlation of individual wage level and annual hours worked, by gender, ages 25–59, United States, United Kingdom, Canada and Sweden, 1967–2005. *Reading note*: The correlation between the annual hours worked and earnings per hour among US males changed from −0.10 in 1967 to +0.10 in 2005. Source*: Blundell and Etheridge (2010)*, *Brzozowski et al. (2010)*, *Domeij and Floden (2010)*, *and Heathcote et al. (2010)*.

between such earners for the United States. It has roughly doubled over 1975–1990, which is less than half the 40-year period, and remained largely stable since. However, the level and evolution of this may differ between countries and, apparently, over time. Conversely, household joint labor supply may also affect the dispersion of wages, if additional earners would operate their labor supply at a less extensive margin of pay or working hours, given that a main income is already secured in the household, or if they would trade off pay and hours for a scenario combining paid labor with other activities, such as household care or participation in education.

In the end, household formation and the two distributions will all be endogenous to each other, and household formation should be added to the list of "stages for comprehending the distribution of income: aggregate factor incomes, differences in earnings and in capital incomes, the role of the corporate sector and of financial institutions, and the distributional impact of the state" (Atkinson, 2007a, 20).

18.2.2 A Cursory Review of the Literature Related to Household Incomes Distribution and LMIs

The literature on the linkage between the two distributions is diverse and cannot be viewed yet as a strong and coherent strand. More than occasionally contributions to

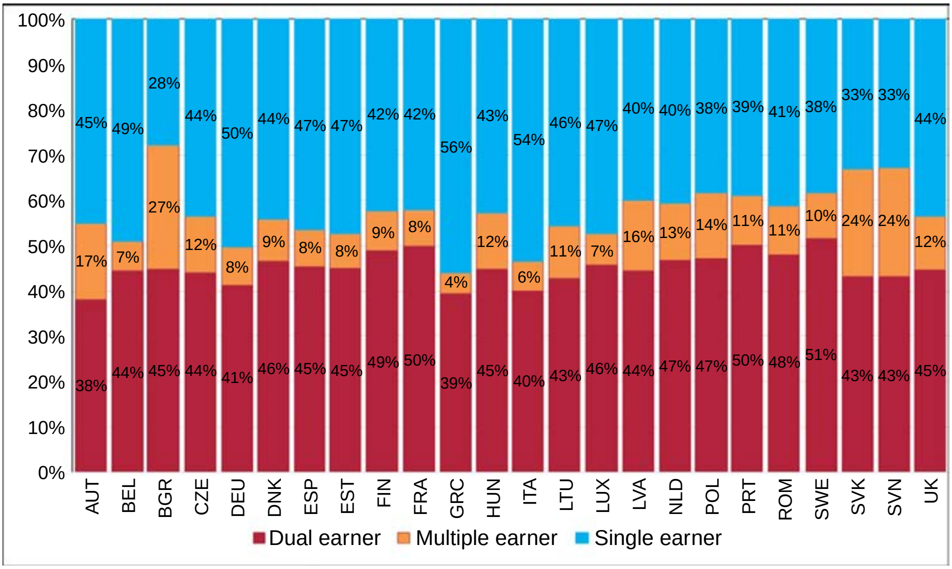


Figure 18.5 Working-age households with employees by number of earners, 26 European countries, 2010. *Reading note*: In Austria 38% of households with at least one member in employment have two persons employed, 17% have three or more persons employed, and 45% have one person employed (including single-person households). *Explanatory note*: Earners need to have positive hours and earnings as well. The household main earner is aged below 65, Students as identified in the data set are excluded. Naturally, female employment participation, traditionally large in what are now former communist countries, is an important determinant. Source*:* Salverda and Haas (2014, *Figure 3.9)*.

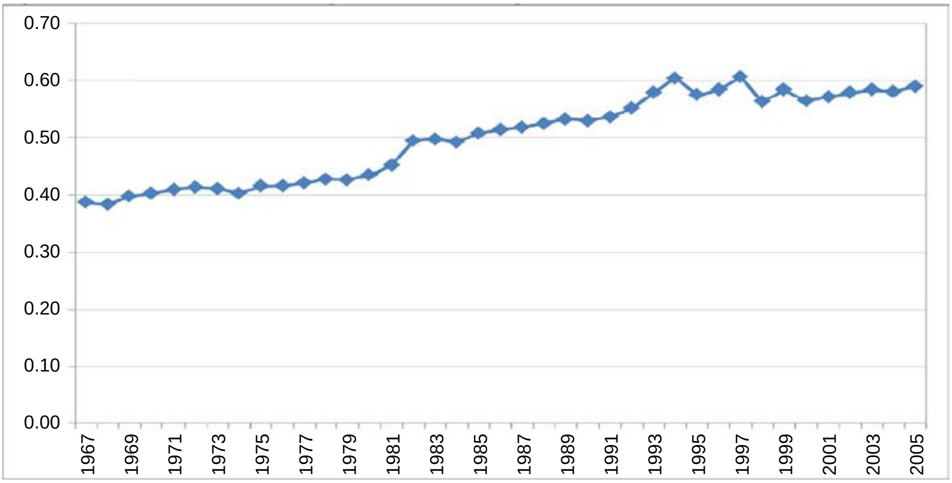


Figure 18.6 Correlation of earnings between married partners, United States, 1967–2005. *Reading note*: Correlation of earnings levels between married partners in households from less than 0.40 to around 0.60. Source*:* Heathcote et al. (2010).

the subject are found in papers dedicated to other issues than the income distribution, such as the design of transfer programs (e.g., Liebman, 1998). Our own reading of the literature on household incomes distribution leads us to conclude that it pays little attention to the role of LMIs, which is, after all, the focus of our chapter. This is the very reason that we only touch on the household context of the dispersion of wages here. Checchi and Garcia Peñalosa (2008, 2010) do address LMIs and income inequality. In a comparative cross-country and macroeconomic perspective, they show the relevance of institutions especially in terms of their effects on the level of unemployment (i.e., zero hours and earnings) which, in turn, contributes significantly to the level of income inequality.[10] We will elaborate on elements of their approach later in the chapter. Certainly, some contributions investigate the effects on the income distribution of one particular institution, the minimum wage—itself the subject of a large literature for its effects on the dispersion of wages. Charles Brown (1999, Section 9.2) in his survey of that literature observes that many families have several earners, so that a minimum-wage worker can be part of a relatively affluent family and adds that the level of the minimum wage will be of little help in reducing income inequality, basing his argument on simple statistics that show the poor fraction among low-wage workers is low and that many poor families have no workers. Neumark and Wascher (2008) sum up many of their own and other contributions to the minimum-wage literature. In their view the combined evidence of income and employment effects for the United States is best summarized as "indicating that an increase in the minimum wage largely results in a redistribution of income among low-income families" (p. 189), as some may see their income rise and others may see their employment and therewith their income diminish. However, Arindrajit Dube (2013) finds sizable minimum-wage elasticities for the bottom quantiles of the equivalized family income distribution and argues from an evaluation of the existing literature, including works by Neumark and Wascher, that the finding is consistent with that.

There is, however, another emerging literature that studies the role of institutions in connection with the household incomes distribution, especially new institutions of relevance such as parental leave, tax credits, including the American EITC or the British WTC, or entitlements to remain in the same job (e.g., Brewer et al., 2006; Dingeldey, 2001; Dupuy and Fernández-Kranz, 2011; Eissa and Hoynes, 2004, 2006; Mandel and Semyonov, 2005; Thévenon, 2013; Thévenon and Solaz, 2013; Vlasblom et al., 2001). However, mostly it preoccupies itself with the employment effects and ignores the income side, and it is strongly focused on particular aspects of inequality as, for example, female labor supply or the motherhood gap in employment participation, and does not consider the aggregate picture of inequality nor the effects on earnings inequality or the

[10] OECD (2011) also advocates including the unemployed zeros in studying the contribution of the earnings distribution to the income distribution.

interrelationship between the two distributions.[11] We leave that literature out here, though we will try in the material to come to incorporate some of those new institutional measures in our broader framework. Note, finally, that we leave out the demographically motivated literature that focuses exclusively on the contribution to income inequality of household structure and composition (e.g., Brandolini and D'Alessio, 2001; Burtless, 2009; Peichl et al., 2010); nevertheless we do include contributions considering this in a broader framework that encompasses earnings inequality (e.g., Burtless, 1999).

In the collection of contributions there seem to be two main approaches (see Table 18.A7 in Appendix D for a summary of the relevant literature). The first approach is based on a direct comparison of the different distributions, and the second approach is based on a decomposition of income inequality that focuses on the sources of income, particularly earnings. The latter shows substantial variation in its choice of the measure of income that is decomposed (mainly established aggregate measures of inequality such as the Gini coefficient, but also newly devised ones such as the "polarization index" designed by Corluy and Vandenbroucke, 2013).[12] More importantly, this literature also varies in the precise technique of decomposition that is applied, which matters as the technique affects the outcome. In the literature there is no single generally accepted way of decomposing, which hampers the establishment of stylized facts.[13] This situation partly motivates the first, comparative approach. In addition to this, it can be observed that the decomposition approach takes one of the two distributions as its starting point and does not consider the effects on the other distribution. Thus it remains unclear when, e.g., growing female employment participation increases household earnings inequality if it also raises individual earnings inequality. We briefly discuss each of the two main approaches.

18.2.2.1 Comparing Distributions

One of the first contributions was made by Gottschalk and Smeeding (1997). They discuss various types of distributions and inequality measures on both the earnings and the income side, but largely in isolation of each other. Their conclusion is that "[b]etter structural models of income distribution and redistribution that can be applied across nations are badly needed. Ideally, an overall framework would simultaneously model the generation of all sources of income (labor income, capital income, private transfers,

[11] Liebman (1998, Table 2) finds a slight increase in the incomes shares of the lowest and the second quintiles in total income in the mid-1990s as a result of EITC; nevertheless, these shares remain well below those obtained 20 years earlier. Note also that Hyslop (2001) and Schwartz (2010) look specifically at the contribution of the association of partners' earnings to inequality on the earnings side.

[12] Note that this considers the distribution of employment over households and not the distribution of employment over pay, occupations, or tasks as discussed in Chapter 5.

[13] Gottschalk and Smeeding (1997, 669) express doubts regarding decompositions and point to the rather different outcomes in the literature. Equally, Gottschalk and Danziger (2005, 249) state that they "do not attempt to decompose the change in family income into its component parts because there are many ways to do so and there is no consensus on the most appropriate decomposition." See Shorrocks (1983) for dire warnings and Kimhi (2011) for a recent critique, but also Cowell and Fiorio (2011) for a possible way out.

public transfers, and all forms of taxation) as well as the formation of income sharing units" (p. 676). That is still a tall order today. In the absence of such a framework, decomposition leaves us with "purely accounting exercises" (p. 668).

Burtless (1999) compares the distributions of annual individual earnings distributions on the one hand and personal equivalized incomes on the other hand for the United States between 1979 and 1996. With the help of simple counterfactual exercises regarding the personal income distribution when holding the levels of earnings inequality constant, he finds that two-thirds of the observed increase in overall income inequality would have occurred leaving only one third for the changes in earnings. Within the latter share he attributes 13% of the increase to the growing correlation between male and female earnings in families. Also the increasing share of single-adult families among the population has contributed because the greater inequality within that group.

Reed and Cancian (2001) also simulate counterfactual distributions for the United States over the period 1969–1999, instead of pursuing a decomposition approach. They argue that this simulation allows using multiple measures of inequality, looking at different points in the distribution, and incorporating changes in the marriage rate. They find that changes in the distribution of female earnings account for most of the growth in family income throughout the distribution and disproportionately more at the bottom, leading to a decrease in inequality. By contrast, changes in male earnings account for over 60% of the growth in the Gini coefficient of the family income distribution.

Gottschalk and Danziger (2005) analyze in an interconnected way the evolution of inequality in four different percentile distributions: hourly individual wage rates, annual individual earnings (and therewith annual hours), annual family earnings, and annual family adjusted total income. The first two distributions are at one side of the earnings–incomes gulf, the other two at the other side. Interestingly, they bridge the gulf by ranking individuals for their annual earnings according to the total earnings of their households (p. 247) using consistent samples of individuals. Earnings exclude the self-employed and the analysis splits throughout between men and women. The focus is the American evolution over the last quarter of the previous century using CPS data.[14]

Atkinson and Brandolini (2006), though, for the most part considering trends in wage dispersion, compare the Gini of the individual annual earnings dispersion to the Gini of adjusted disposable household income for a set of eight countries: Canada, Finland, Germany, the Netherlands, Norway, Sweden, the United Kingdom, and the United

[14] Gottschalk and Danziger's approach is very apt in an intertemporal perspective but difficult to interpret in a cross section as it ranks male and female earners according to their respective households, which must be largely overlapping sets that concentrate higher up the income distribution, to the extent that both male and female in a household do have earnings. A disadvantage is that they do not discuss the role of singles nor of possible third earners within the household. They find that "for females, changes in hours more than offset the rise in wage inequality. The acceleration in male wage and earnings inequality during the early 1980s disappears when earnings of other family members are included" (p. 253). Thus, the household is found to mitigate inequality growth in the labor market.

States, using LIS data from around the year 2000. They draw the comparison on an annual basis and include part-time and part-year earnings, but they leave the distribution of employment out from their analysis, and, consequently, they also do not compare directly to the hourly wage rates, the traditional pay inequality in the labor market. In addition, they do not compare individuals and households on the basis of an identical ranking as is done by Gottschalk and Danziger. They find that the Nordic and Continental countries have similar Gini values for earnings and for incomes respectively, whereas both are higher for Canada and the United States; the United Kingdom is found to be European on earnings and North American on incomes (p. 58).

Lane Kenworthy (2008) observes that "if every household had one employed person, the distribution of earnings among households would be determined solely by the distribution of earnings among employed individuals" (p. 9). He mentions the possibility that households have different numbers of earners, adding that this number is mainly determined by the number of adults in the household. However, he leaves this aside in the analysis and focuses on the dichotomy between "some earner(s) or none" (p. 9). Using LIS data for 12 countries (Australia, Canada, Denmark, Finland, France, Germany, Italy, the Netherlands, Norway, Sweden, the United Kingdom, and the United States), he finds pretaxed, pretransfer household income inequality to be strongly related to the inequality in individual earnings of full-time employed individuals, all equivalized for household size and composition. The association to the incidence of households with zero earnings (for the head of household) is less, and to marital homogamy, defined as the correlation between spouses' annual earnings, it is smaller still. The total employment rate and the part-time employment rate appear to play no role.

Večerník (2010), also using LIS data, considers employees only and does so in conjunction with their households. His focus is the effects of transition in four CEE countries, in a comparison with Germany and Austria. He specifically draws other earners than the spouses in a household into the comparison, and effectively distinguishes between dual-earner and multiple-earner households. He shows that the latter category of employees can make an important contribution to household earnings, that earnings inequality among this group is very high in all countries, and that the contribution to overall inequality can also be very substantial. Slovakia combines the highest earnings share (19%) with a lower Gini coefficient than elsewhere, and a major contribution to overall inequality (39%). This contrasts strongly with Germany where both the income share and the contribution to overall inequality are the lowest (4% and 8%) and the within-group inequality is the highest (0.93). It seems to suggest that the population of other earners may have a very different character in Western Europe than in the East.[15]

[15] Večerník (2013), studying the evolution of the two distributions in the Czech Republic between 1988 and 2009, again with the help of regressions on both sides, finds an important role on both sides for education which runs via the employment and earnings of women as marital partners.

Finally, Salverda and Haas (2014), using EU-SILC data, build on some of the above approaches comparing decile distributions and the top-to-bottom inequality ratios (the shares or means of the tenth top decile relative to that of the first decile) in a cross section of 25 EU countries in 2010. They show how the dual-earner households and especially the multiple-earner households concentrate toward the top of the household earnings distribution: on average, across EU countries only one-tenth of households in the top decile are single-earner households whereas almost 90% are in the bottom decile (compare Figure 18.5 for the average picture). Unsurprisingly, dual-earner and multiple-earner households reach the top by combining wage levels often from well below the top of the earnings distribution, in contrast to the few single-earners whose households make it to the top. On average over the countries, the main earner's earnings are only 60% of a single earner's in a dual-earner household and less than 50% in a multiple-earner household. Salverda and Haas draw a comparison of the household earnings distribution with two different ways of distributing the individual earners: one ranked according to their households' earnings, the other ranked by their own individual earnings. They find that households add to household earnings inequality primarily by the combination of the activities of their members, although at the same time that combination mitigates the individual labor market inequalities in both hours worked and levels of pay: workers with higher earnings or longer hours combine with those working or earning less. At the same time, in international comparison the variation in hours is modest—clearly, one can only work so many hours regardless of the country—and the main difference reflected in the comparative level of household earnings inequality is, after all, the traditional inequality of the individual's own wages in the labor market.

Figure 18.7 compares household total earnings to individual wages in panel (a), and to hours worked in panel (b). The lower level of individual earnings inequality and annual-hours that is attained if persons are ranked by their households (lines 3 and 6) instead of as individuals in the way they appear in the labor market (lines 2 and 5), shows the mitigating effects of households compared to the labor market. Households earnings and hours (lines 1 and 4) are more unequally distributed due to the adding up of individual earnings, which, however, are attained at lower and higher levels. When compared to panel (a), panel (b) also shows that the inequalities in hours are substantially smaller than in earnings within as well as across countries. This is understandable as there are only so many hours in a year and the number of employees combined in a household is modest in practice.

18.2.2.2 Decompositions of Household Income Inequality

The second relevant approach in the literature is based on decompositions of income inequality, especially by sources of income which enables scrutinising the contribution that earnings or employment make to inequality. There is significant variation among the decomposition studies: their nature and the variable decomposed, and also the

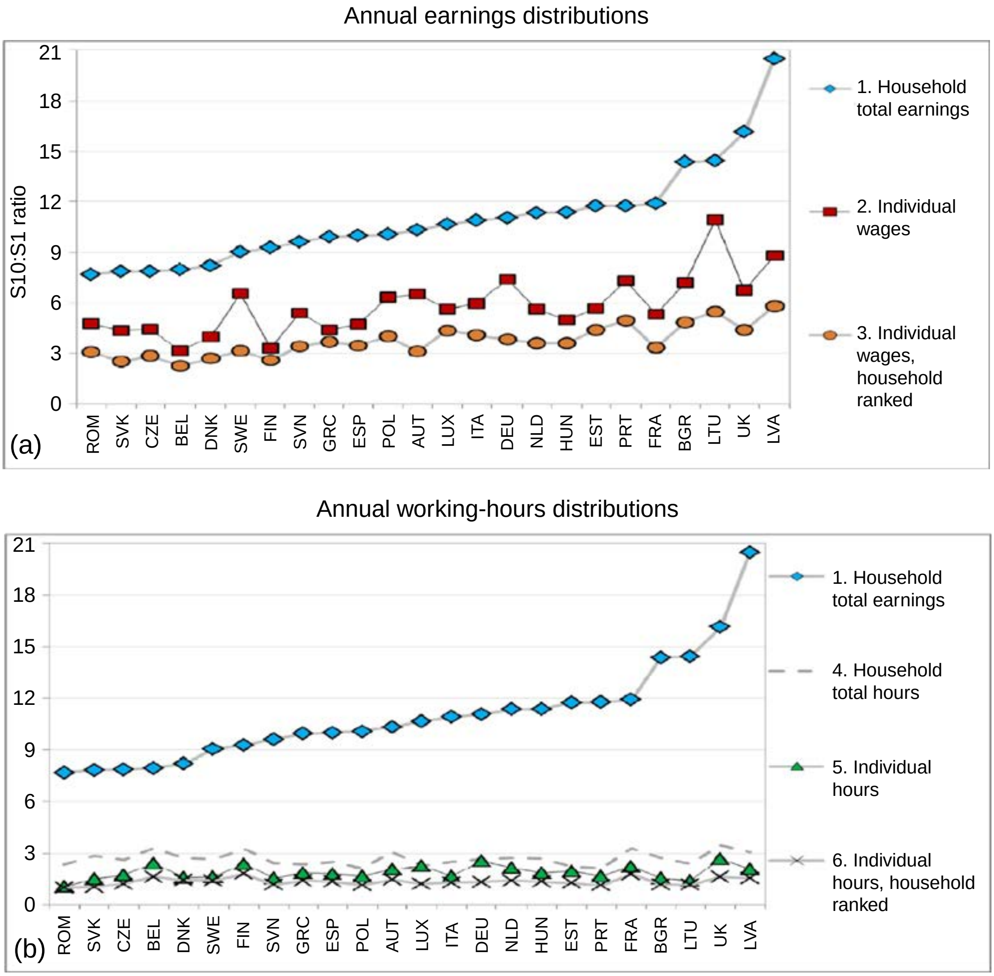


Figure 18.7 Top-to-bottom ratios (S10:S1) for employed individuals and their households, 26 European countries, 2010. (a) Annual earnings distributions. (b) Annual working-hours distributions. *Reading note*: In Romania average household total annual earnings in the 10th decile of such earnings are 8 times higher than in the 1st decile; annual earnings of individuals in the 10th decile of such earnings are 5 times higher than in the 1st decile if persons are individually ranked, and only 3 times if they are ranked according to their households total annual earnings. *Explanatory note*: The sample concerns households receiving their main income from earnings. The top-to-bottom ratio is between the average level of the tenth decile to the first decile. Source*: Salverda and Haas (2014, Table 3.2).*

technique of decomposition (see Fortin et al., 2011 for an overview). The results may depend on the choice.

In one of the first studies, Shorrocks (1983) using the American PSID over 1968–1977 concludes that "Dollar for dollar capital income and taxes have more distributional impact than earnings, which in turn exceeds the impact of transfer income" (which is defined to include retirement pensions and annuities).

Van Weeren and Van Praag (1983) use a special data set covering seven European countries (Belgium, Denmark, France, [West] Germany, Italy, the Netherlands, and the UK) in 1979 to decompose income inequality into subgroups. Interestingly, they look, inter alia, at the employment status of the head of household as well as the number of persons contributing to household income. At the time both characteristics make the largest contribution to inequality in Denmark, although employment makes the smallest contribution in the Netherlands and the number of earners in the UK.

Blackburn and Bloom (1987) draw a careful comparison of the family annual earnings distribution and the individual annual earnings distribution for the United States over the years 1967–1985. Using various aggregate inequality measures they find that annual earnings inequality has hardly changed, although income inequality has. Descriptively splitting the distribution in five parts, the change seems largely concentrated in what they term the "upper class," family with earnings over and above 225% of the median. From a time-series regression analysis they conclude that particularly the growth of nonprincipal earners in those households contributes to this growth. Blackburn and Bloom (1995) draw an international comparison at various points during the 1980s. For the United States, Canada, and Australia they find that income inequality increased among married-couple families and that the increases are closely associated with increases in the inequality of husbands' earnings. Evidence of an increase in married-couple income inequality is found also for France and the United Kingdom, but not for Sweden or the Netherlands. In various countries, that increased inequality of family income is closely associated with an increased correlation between husbands' and wives' earnings. A more detailed examination in Canada and the United States suggests that this increase cannot be explained by an increase in the similarity of husbands' and wives' observable labor market characteristics in either country. Rather, it is explained partly by changes in the inter-spousal correlation between unobservable factors that influence labor market outcomes.

Karoly and Burtless (1995) follow Lerman and Yitzhaki (1984) in decomposing the evolution of the Gini coefficient of American distribution of personal equivalized incomes between 1959 and 1989, basing themselves on census and CPS data. They find largely the same results as Burtless (1999) does for his more recent period. A large part of the reduction in income inequality before 1969 is attributed to the decline in earnings inequality among male heads of families. After 1969 the same group is responsible for more than one-third of the increase in inequality. Since 1979, the improved earnings of women have increased inequality as they were concentrated in families with high incomes.

Cancian and Schoeni (1998) consider 10 countries using LIS data for the 1980s. They find that the labor-force participation of wives married to high-earning husbands increased more than for those married to middle-earning men.[16] At the same time,

[16] They do not decompose strictly speaking but use a simple split of the coefficient of variation between married partners to look at the contribution of wives to inequality among this category; they therefore do not address the income distribution as a whole.

the mitigating effect of wives' earnings actually increased slightly in all countries. In their view an unprecedented increase in the correlation of earnings between the partners would be needed to make the effect disequalizing.

Evelyn Lehrer (2000) finds from the US National Survey of Families and Households that between 1973 and 1992–1994 the equalizing influence of the wife's contribution grew substantially stronger—partly due to a decrease in the dispersion of female earnings relative to that of male earnings. This seems to contrast with Karoly and Burtless (1995); however, her finding relates to married couples and their earnings only, not to the full personal income distribution.

Del Boca and Pasqua (2003) consider husbands and wives in Italy between 1977 and 1998 using regional differences and the absence of wives' incomes as a counterfactual. The added worker effect is found in households especially in the North where there is more acceptance and more choice of working hours and more child care support available. Here the reduction in the dispersion of wives' earnings seems to have offset increases in the dispersion of husbands' earnings as well as the increased correlation in the earnings between the spouses between 1989 and 1998.

Johnson and Wilkins (2003), following DiNardo et al. (1996), studying Australian inequality over the period 1975–1999, find changes in the distribution of work across families—for example, an increase in both two-earner families and no-earner families—were the single-most important source of the increase in private-income inequality, with such changes on their own accounting for half the increase in inequality.

Daly and Valetta (2006), using CPS data for the United States and adopting partly the method of Burtless (1999), in combination with the decomposition technique proposed by DiNardo et al. (1996), find a more substantial contribution (50–80%) of men's earnings to increased American inequality between 1969 and 1989 than does Burtless. This increase was counteracted by the growing employment participation of women. They explain the larger role of males as their methodology can account for growing inactivity and unemployment.

The *Review of Economic Dynamics*' Special Issue of 2010[17] presents an interesting and important inventory of various dimensions of economic inequality, including the distributions on both sides of the individual earnings versus household incomes divide as well as the distributions of wages versus that of hours. The set of papers for seven countries contains useful descriptives of the distributions. In addition, some decomposition

[17] Relevant to the set of countries covered here are Canada: Brzozowski et al. (2010), Germany: Fuchs-Schündeln et al. (2010), Italy: Japelli and Pistaferri (2010), Spain: Pijoan-Mas and Sánchez-Marcos (2010), Sweden: Domeij and Floden (2010), UK: Blundell and Etherigde (2010), and finally United States: Heathcote et al. (2010). In spite of the fully comparative set-up from the start, there are still some incomparableness left, especially with regard to annual individual earnings and to the household earnings distribution which is not always given on the same basis (pregovernment, pretax, after-tax or equivalized disposable income).

exercises are done on the log-variance of either earnings or hours. These decompositions concern a limited but important range of characteristics (gender, education, age, experience, region, family structure). They appear to explain little of the evolution and, in virtually all cases, leave most of the action to the residual. Of particular interest is Figure 18.8, where panel (a) specifies the variance of log individual hourly wages and panel (b) that of individual annual hours worked. The two are at different levels, the latter nowadays being much lower than the former, and their evolution seems to trend in opposite directions, clearly up for the former and declining for the latter. For annual earnings—seldom known from the contributions—the implication is a more substantial variance, which then feeds into household earnings.

Lu et al. (2011) study Canadian developments in the family earnings distribution (equivalized) from 1980 to 2005 using census data for those 2 years and 1995. They again adopt the decomposition approach developed by DiNardo et al. (1996). For 1980–1995 they find substantial increases in family earnings inequality, but for 1995–2005 some decrease. Changes in the earnings structure, such as those attributed to educational attainment, and changes in family composition (fewer married couples, more single individuals and lone parents) have been key factors contributing to growing family earnings inequality. Substantial changes in family characteristics (including a surprising decline in educational homogamy and the implied mating of women below their level) have had the most important counteracting effects as has continued growth in women's employment rates. Interestingly, the authors take a special look at the Top 1% of the distribution, mention that it has increased substantially between 1995 and 2005 in contrast with declining family earnings inequality; however, they do not further highlight this in their analysis.

Larrimore (2013), again focusing on American CPS data, now for 1979–2007, and with the help of a shift-share decomposition, finds important differences between the three subsequent decades: changes in the correlation of spouses' earnings accounted for income inequality growth in the 1980s but not in the 1990s (consistent with Figure 18.6). During the 2000s changes in the earnings of male household heads diminished income inequality, and the continued growth in income inequality was due to growing female earnings inequality and declining employment of both genders.

Finally, the most extensive decomposition study seems to be the one reported by Brewer et al. (2009) and Brewer and Wren-Lewis (2012). For the National Equality Panel (Hills et al., 2010) they have dissected the trends in British inequality over the long period 1968–2006 in many respects using a regression-based decomposition technique developed by Fields (2003) and Yun (2006).[18] The results are presented in Figure 18.9. Total inequality of all households (line with white markers) moves to a higher level over the 1980s, from less than 100 to more than 160. The contribution that

[18] Unfortunately they compare gross earnings to equivalised disposable household incomes, but they do decompose between (aggregate) taxes and benefits.

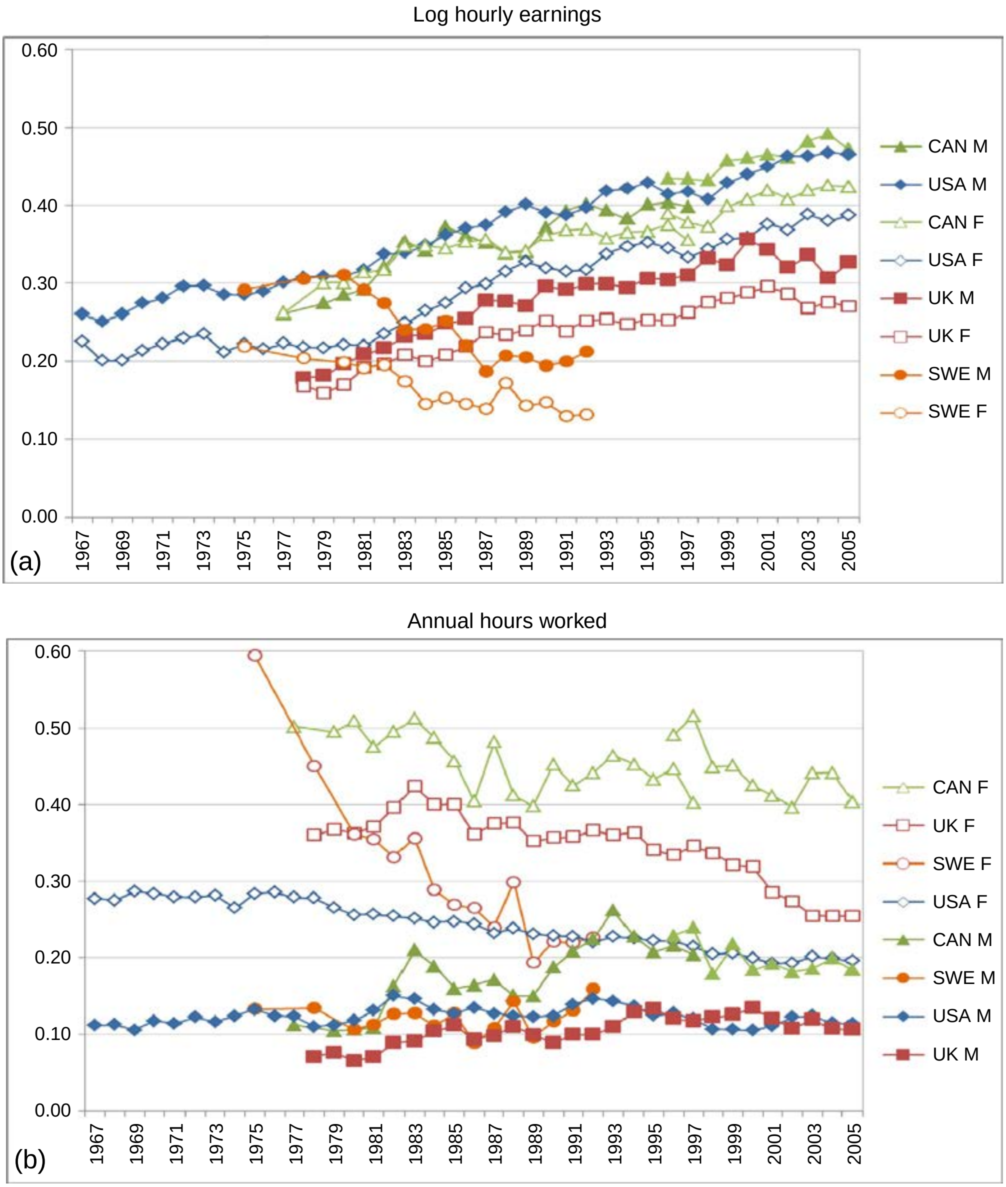


Figure 18.8 Evolution of variance, by gender, ages 25 to 59, United States, United Kingdom, Canada and Sweden, 2010. (a) Log hourly earnings. (b) Annual hours worked. *Reading note*: The variance of log hourly earnings of US males increased from 0.26 in 1967 to 0.47 in 2005. *Explanatory note*: F—females, M—males. Source*: Blundell and Etheridge (2010)*, *Brzozowski et al. (2010)*, *Domeij and Floden (2010)*, *and Heathcote et al. (2010)*.

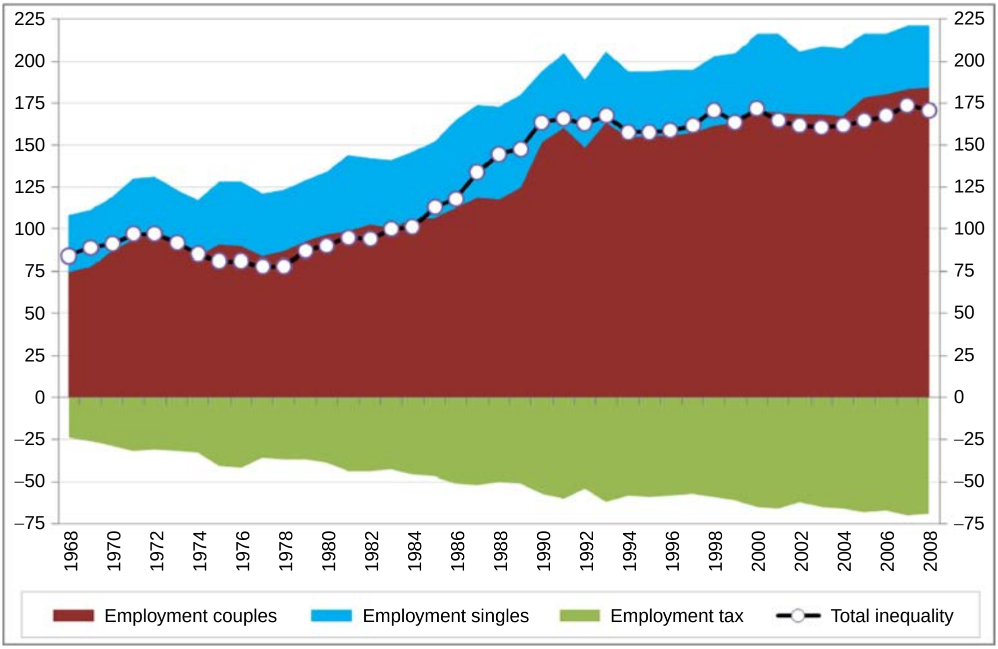


Figure 18.9 Contributions of household earnings to total net-equivalized household income inequality, United Kingdom, 1968–2008. *Reading note*: The line of total inequality results from adding up the contributions to inequality from couples and singles in employment and subtracting the tax they pay. *Explanatory note*: Inequalities are measured as the variance of logs (×1000). Contributions do not exactly add up as nonemployee categories receiving market income, pension have been left out. These contributions happen to partly cancel out but their aggregate has grown from 0 points in 1968 to 19 points out of the total of 171 that is shown for 2008. Source: *Brewer and Wren-Lewis (2012, Table 5).*

household gross earnings makes to this is split between the single-earner and dual-earner households respectively and the total of taxes paid by both (stacked shaded areas). The role of singles has remained unchanged on balance, with a temporary increase during the 1980s. Dual earners run largely parallel to total inequality; their growth is also somewhat concentrated to the 1980s though it continued after that at a slower pace. Taken together single and dual earners lag the inequality growth of the 1980s somewhat. That gap is filled by incomes from self-employment, investment, and pensions whose role more than doubled during the 1980s (not shown).[19] The net effect of earnings is less as taxation (the negative area which needs to be deducted) has also increased. After an

[19] The relative role of benefits (including tax credits) grows until the mid-1980s but is almost halved subsequently.

initial rise up to the mid-1970s the rise is more gradual and extends over the period as a whole but hardly changes relative to earnings.

At the end of this overview a careful and detailed comparison of these results, including replication studies, seems advisable to find out where they diverge or even contradict and to seek an explanation whether differences are real—i.e., related to the period or the sample that is the focus—or artificial—i.e., due to the data set, the method of decomposition, or the approach to equivalization. Unfortunately, however useful, such a meta-analysis is entirely outside the scope of our contribution.

18.2.2.3 A Heuristic Help for the Role of Institutions and Earnings

Though we cannot and will not pursue a comprehensive approach to wage dispersion and income distribution, we may still ask what we can learn from the above and take with us for the contemplation of wage dispersion and institutions. We need to keep in mind, first and foremost, that labor market earnings make a major contribution to household incomes as well as their dispersion. By implication, the lack of such earnings resulting from unemployment or joblessness makes a large contribution, too.

Important developments are found that tend to diminish the direct influence of wage dispersion on the income distribution as the growing female labor market participation and at the same time enhance the role of household joint labor supply. This complicates the relationship between the two distributions, and it may also affect the labor market behavior of labor supply. Anyway, it brings into play a collection of new institutions that may affect both employment, hours worked, and pay, as well as their concentration across households. This may influence the level of wage inequality. It seems advisable to take the new institutions into account in addition to the traditional ones arising from labor market analysis on its own.

Another important inference to draw is the importance of considering hours and their dispersion in addition to wages. The inclusion of hours is important for several reasons. They are needed to arrive at the full picture of the earnings input that the labor market makes into household incomes. The hours dispersion differs significantly between the sexes, between countries, and also changes over time. In addition, the growing role of part-time and temporary jobs in itself makes this a more important dimension, and one that may also play a role in determining the dispersion of pay given the correlation between hours and pay. There may also be different trade-offs between hours and pay in different countries. At the same time, the role of hours may be relatively less important; it is more modest because of natural constraints than that of pay in an international comparison.

Second, it seems safe to conclude that one size does not fit all (countries). Significant differences are found, especially between different periods, and these seem to get more attention the further behind the period is (witness Larrimore's most detailed account of such periodization in his 2013 publication).

Interestingly, comparable decompositions of important characteristics such as gender, age, education, and family type seem to play an amazingly small and also often flat role in virtually all countries, leaving a large role to residuals, which may point to national idiosyncrasies.

18.3. WAGE DISPERSION: MEASUREMENT AND STYLIZED FACTS

Before we turn to the analysis of wage inequality and institutions in Section 18.4 we discuss here first the ways to measure these and then present what seem to be the current stylized facts of the literature concerning wage inequality. Section 18.3.1 starts with a discussion of the issues involved in measuring wage inequality and a quick presentation of data sources. This is followed by a presentation of the "stylized facts," which we define as the state-of-the-art knowledge of wage inequality currently accepted by scholars as necessitating explanation in spite of their different views and approaches. These facts regard, first, the aggregate level of inequality, referring to the most comprehensive distribution at the national level. For this we discuss outcomes according to different measures of inequality as well as for different definitions of the wage variable. Second, Section 18.3.2 considers disaggregate inequality, which highlights specific parts of the distribution—such as the tails or the middle—on the one hand, and inequalities among various subsamples of the population according to demographic or labor market criteria on the other hand. Then (in Section 18.3.3) we provide some new empirical evidence from a cross-section comparison of 30 countries for the most recent year available, which we elaborate on when our empirical approach in Section 18.3.4 concludes.

18.3.1 Measuring Wage Inequality and Data Sources

Blackburn and Bloom (1987) have argued in detail the need of precision for measures and definitions of wage inequality.[20] Following their suggestions, we need to pay attention to at least four dimensions:

[20] "The often-contradictory conclusions reached by studies of recent trends in income and earnings inequality are largely explained by the reliance of researchers on a remarkably wide range of conventions of data analysis. For example, the list of important dimensions in which previous studies vary includes: the time period covered; the way family units are defined; the population to which the studies of individual earnings generalize (e.g., all earners, private nonagricultural workers, male earners, wage and salary workers, full-time, year-round workers, etc.); the measures of earnings and income (e.g., total family income, equivalent family income, total family earnings, wage and salary income, etc.); the unit of time for the measurement of earnings (e.g., annual, weekly, or hourly); the nature of the earnings measure (e.g., usual earnings or average earnings); measures of inequality (e.g., the Gini coefficient, income-class shares, variance of logarithms, coefficient of variation, mean logarithmic deviation, etc.); the use of individual or grouped income/earnings data; the treatment of sample weights; the treatment of observations with imputed incomes; the handling of top-coded values of income and earnings; and other criteria for including observations in the sample, such as the age of the respondent and whether the respondent was working at the time of the survey or in the year preceding the survey" Blackburn and Bloom (1987, 603).

(1) the measure of inequality
(2) the definition of the wage variable (including its time dimension)
(3) the selection of the sample of the population that is being covered
(4) the nature of the data sources.

Clearly, the study of wage inequality adds several significant issues of measurement to those of long-term concern to the study of inequality (e.g., Atkinson, 1970; Chapter 5; Jenkins and Van Kerm, 2009). We consecutively address these four issues before we turn to data sources, and to the stylized facts in the following section.

Before starting this we mention a general observation. Wages are defined here as "wage rates,"[21] preferably controlled for hours worked[22] and therewith for differences in workers' efforts, whereas we consider "earnings" or "wage earnings" as the product of those wage rates with the hours worked and therefore reflecting also differences in individual efforts. For convenience we say in general that we are addressing "wage inequality." However, this does not mean that we restrict ourselves to the inequality of wages rates only; to the contrary, we aim to also consider the dispersions of hours and earnings.When doing so we will try to be clear and not just mention wages but use the appropriate concepts: weekly, monthly, or annual hours or earnings.[23] Wage rates serve the clear analytical purpose of enabling comparisons between individuals on the basis of the same efforts made in terms of time dedicated to paid work, measured in hours. As already argued, hours are an increasingly important dimension of labor market functioning and inequality and will be given their due.

18.3.1.1 Measures of Inequality

Although the Gini coefficient is a very popular measure in the analysis of income inequality, it hardly figures in the analysis of wage inequality. Variance, mean log deviation, the Theil index, and standard deviation are used, however.[24] Unfortunately, because of their aggregate nature, these measures tell us little about where in the distribution the differences over time or across countries reside, though decomposition of these measures, as far

[21] "Wage rates" as hourly wages can be part of wage scales agreed between unions and employers, albeit implicitly, when the agreement also covers hours of work. However, actual individual earnings will often deviate from these scales because of bonuses, performance pay, labor market scarcities, etc. (see, e.g., Salverda, 2009).

[22] Note that this may add to measurement error.

[23] Here we differ from OECD (2011, 26), which follows a more complex scheme that risks creating confusion: Their "dispersion of hourly wages" equates to our dispersion of wages and their "wage dispersion" equates to our "distribution of annual earnings." "Labor income" is a concept encountered in the US inequality literature and is effectively considered as a wage rate; however, it actually amounts to a wage rate multiplied by the efforts (usually for full-time workers on a weekly basis).

[24] There is an extensive literature discussing the properties and validity of these measures, such as the violation by the standard deviation of the transfer principle—see, for example, Chapter 6 of this handbook or Jenkins and Van Kerm (2009). Compare, e.g., Karoly (1992) who considers empirical outcomes for a broad range of such measures for American wage inequality.

as possible, can certainly be helpful for understanding the underlying processes. In wage-inequality analysis it is the percentile ratios that play a remarkably important role: the P90:P10, P90:P50, and P50:P10 ratios, which mutually relate the 10th, 50th, and 90th percentiles to each other.[25] These ratios are directly helpful in focusing attention on particular parts of the wage distribution and they are intuitive at the same time. Their evolution over time reflects differential changes in wages at specific points of the distribution. As we will see below, up to this very day the debate on the effects of the minimum wage on wage inequality is framed almost exclusively in terms of these ratios. The ratios have also provided important leverage to the shift that has occurred in the debate about the role of technology as a determinant of growing wage inequality. Their popularity may relate also to an easier consistency with the analytical focus on the individual and his or her efforts in the labor market in contrast to income analysis.[26] Note that the ratios are based on the upper-boundary wage levels of the chosen percentiles (or deciles), and not on their means, sums, or shares in the total of wages. This implies certain limitations to the use of these ratios, and it seems advisable to add measures that broaden to averages, sums, or shares. For example, a top-to-bottom ratio between the means, sums, or shares of the top decile on the one hand and that of the bottom decile on the other hand (denoted as S10:S1) may find inequality growing much farther apart than the P90:P10 ratio would suggest, if important changes are actually occurring within the two tail deciles and affecting their within-spread.[27] Precisely that is the upshot of the recent analysis of top-income shares, where the sum and the share of the top decile, and its within-distribution over smaller fractions, are the very subject of study. In a similar vein, much of the current minimum-wage debate appears to be effectively analyzing changes found within (and perhaps even restricted to) the bottom decile of the wage distribution. Note that the OECD has recently introduced the top-to-bottom ratio in its income inequality and poverty database.[28] In addition to these quantile ratios, the ratio between the average wage and the median wage is sometimes also found as an indicator of wage inequality; the Kaitz index similarly relates the level of the minimum wage to the average wage in the analysis of minimum wage effects. One disadvantage of all such ratios, however, is that they cannot be decomposed (Lemieux, 2008, 23),[29] though they may be further split into ever smaller fractions.

[25] Also denoted as decile ratios: D9:D1, D9:D5, and D5:D1, between the 1st, 5th, and 9th deciles, the aforementioned percentiles being their upper boundaries. Comparisons for all percentiles encountered in the literature below may be considered a visual generalization of this type of measures.

[26] Relative to the individual employee the type of the household as a unit of analysis shows much more variation, which is difficult to square with the use of exact percentile income levels as it is accidental what type may be found at a particular income level.

[27] Compare the "poverty gap," which acts as an indicator of the within-spread of poverty.

[28] See http://www.oecd.org/els/soc/income-distribution-database.htm.

[29] However, Firpo et al. (2009) develop a decomposition method based on recentered influence function (RIF) regressions, which they actually apply to these ratios.

Some other indicators are available in the same family of disaggregate measures that can also relay information about wage inequality. These relate to parts of the distribution that are defined with the help of an external wage-level criterion. The most important one in practice, regularly published by the OECD, is the incidence of low-wage employment (see Gautié and Schmitt, 2010; Lucifora and Salverda, 2009). This is defined as the share of all employees in the wage distribution who are found having wages below the level of two-thirds of the median wage.[30] It is important to realize that this is a concept that relates to the analysis of the labor market, in contrast with in-work poverty that depends on the household-income position of the wage earners concerned; nonetheless the former is definitely relevant to the analysis of the latter. The concept of low pay is only infrequently used in US analyses of wage inequality where the in-work poverty concept is more frequent, perhaps because the poverty threshold is of such central concern in that country's public discourse.[31] The divergence between the two concepts signifies that workers may be poor—on the basis of their household situation—at wage levels that are well above the low-pay threshold, and vice versa, that workers receiving low pay may be found in households well above the poverty level.[32] Unsurprisingly, the evolution over time may differ between low pay and poverty wages. Figure 18.10 clearly points this out for the United States. Over the period 1995–2002 the share of employees earning poverty wages shows a particularly sharp decline, although the incidence of low pay remains unchanged. Household composition, household joint labor supply, and the evolution of prices determining the poverty lines can influence the former but not the latter, which depends on wage developments.

As an analogue to low-wage employment one can conceive of the incidence of pay at or below the minimum wage as another simple measure of wage inequality. Strikingly, in spite of decades of intense debate on the employment effects of the minimum wage such

[30] Although there is a clear and internationally endorsed measure of low pay this is not readily available for high pay. Salverda et al. (2001) define high pay as over and above 1.5 the median wage, but other definitions are also found in the literature. The OECD Earnings Database also specifies high pay—using the same definition—but so far only for a few countries. By implication, as long as the tails are not well defined there is also no clear measure available of the polarization of the wage distribution, which might easily be defined as what remains in the middle of the distribution after excluding low-wage and high-wage employment. Instead, polarization seems to be gauged more as a qualitative phenomenon from ad hoc visual inspections of real wage growth, as we will see later.

[31] We disregard the debate about the "experimental poverty measures": in principle, the same difference of focus attaches to the European concept of poverty (see Chapters 3, 8, 9, and 23).

[32] For 2011, the US low-pay threshold can be put at $11.89 per hour (EPI *State of Working America 2012*, Table 4C), which at 2000 h of work in a year would generate annual earnings of $21,340, well above the official poverty threshold for a single-person household ($11,702, <65 years) and only slightly below that for two-adult, two-children households ($22,811). The poverty thresholds range up to $50,059, depending on household size and composition, which is 2.3 times low-wage annual earnings. We disregard for a moment taxes and contributions and also that the poverty levels are rather low as underlined by the introduction of the Experimental and Supplemental Poverty Measures.

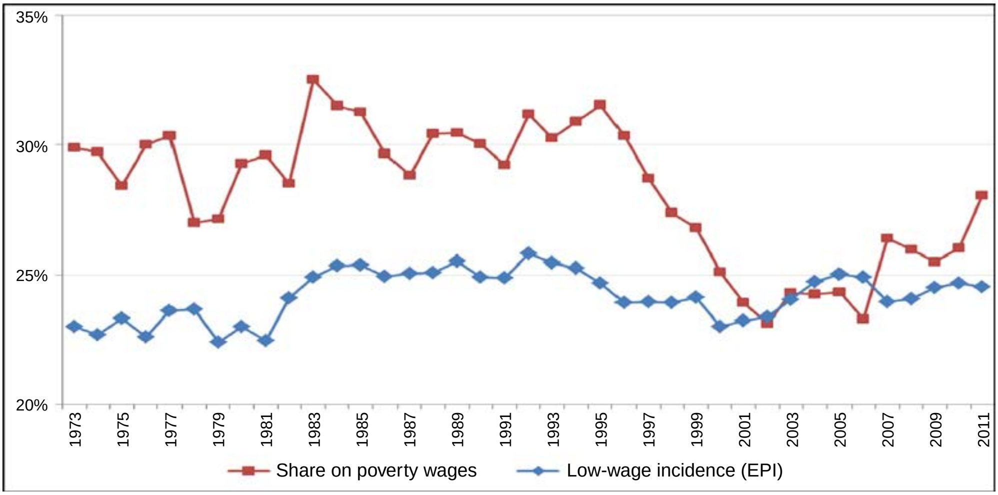


Figure 18.10 Shares (%) of workers earning a poverty wage or a low wage, United States, 1973–2011. *Reading note*: The percentage of all employees earning a poverty wage fluctuates around 30 until 1996 and then falls substantially; the percentage earning a low wage fluctuates around 25 from 1983 onward. *Explanatory note*: Poverty wages are earned by individuals whose household incomes are below the official poverty threshold; low wages are defined as being at or below two-thirds of the median wage: on authors' estimation for hourly earnings of all workers using linear interpolation in the decile distribution. Source*: Authors' calculation on EPI,* State of Working America 2012*, data underlying Figure 4E and C.*

statistical data are sporadic. Internationally, a possible explanation may be the nonuniversality of statutory minimum wages or their complex nature when, for example, it is less evident to whom they apply or not—a problem that is absent in measuring the low pay incidence.

Finally, as implictly suggested above, the share of top wages in the wage distribution—a direct corrollary to top-income shares—provides another possible statistic that can throw light on wage inequality. We will see later that pay at the top plays an increasingly important role in the wage-inequality debate.

18.3.1.2 Definitions of the Wage Variable[33]

Most of the literature restricts the definition of the wage variable to the payments received by employees from their employers, and we will follow that convention here. This excludes for reasons of principle both the unemployed and the self-employed (however, this does not mean that they should be excluded from the analysis of labor markets and wage inequality—compare our approach in Section 18.5). We will focus on gross

[33] For a deeper discussion of these definitory issues and the issues of composition and statistical observation considered next see Atkinson (2008), Chapter 2: *Taking Data Seriously: Where the Data Come From and How We Should Use Them.*

wages, including taxes and contributions, which are paid by the employee (also when the employer actually withholds them on behalf of the tax authorities). However, gross wages are not available for all countries all of the time though, fortunately, they now increasingly are (e.g., very recently France, Greece, and Switzerland started to provide gross wages; net wages will likely show a lower level of inequality because of tax progression). In addition, even gross wages are a more restricted concept than "employee compensation" in the sense that they exclude employer contributions such as for occupational pensions and other provisions. This is for the practical reason of lacking observations in most countries.[34] The full-gross wage defined as employee compensation including employee taxes and contributions seems the most appropriate concept in principle as it includes what can be called the "social wage." This encompasses entitlements financed out of employee and employer contributions and income tax, and varies significantly between countries (Gautié and Schmitt, 2010). Finally, the wage concept mostly comprises payments that are actually made by the employer and may leave out informal cash payments such as tips, in spite of their (suggested but often statistically unknown) importance for low-wage earners in some countries.

Given this definition of wages, there is one crucial dimension about which we aim to be as clear as possible. This regards their time dimension, which appears to greatly influence the apparent level of inequality. We have already touched on this above when mentioning the distinction between hourly wage rates and their multiplication by hours worked. Most of the US inequality debate has been framed in terms of full-time weekly wages if not full-time full-year wages (Acemoglu and Autor, 2011, 1049)—"earnings" in our definition. Though this seems largely a matter of data convenience, it may have important implications for comparisons. First, it ignores the incidence of part-time employment which varies significantly over time and across countries. Second, it overlooks the dispersion of full-time working hours itself, which can be considerable and may differ between countries.[35]

[34] Commonly, such contributions are not well known to the employee, and they are left out in household surveys as a consequence. They may differ considerably over the wage distribution and between countries. Among the stylized facts below, we will, however, mention an excellent example of information on the distribution of employee compensation.

[35] The OECD database on "Usual hours worked by weekly hours band," covering 28 countries, indicates that for 2012, on average, 76% of men and 65% of women worked 40 h or more in 2012, with a highly comparable cross-country pattern for the two genders. However, these shares vary from around 10% only in Denmark to almost 100% in Estonia, Greece, Hungary, Poland, and Slovenia. The evolution over time also differs. On average for the 13 countries with data for both 2012 and 1983, the share among male full-timers decreases from 81% to 66%, among women from 69% to 53%. In various European countries the shares working longer hours plummets: for example, in Denmark from 95% (men) and 85% (women) to 12% and 8%, respectively, in Germany from 100% and 99% to 73% and 64%, respectively. However, in the United States it remains unchanged at slightly above 90% and the growing number of women adapts upward to males. This female adaptation is also found in the United Kingdom, but not in various other countries, where the gap can even grow.

Third, different time periods for wages/earnings bring into play different, additional elements of pay such as bonuses and other special payments that are made with a lower frequency, for example, on an annual basis. Such payments usually have an increasing effect on inequality, which risks to be missed by a shorter time horizon—the use of an annual average of shorter-term wages can potentially mend this problem, but this is not standard practice.

Fourth, the use of time for the weighting of the observations bears on the level of inequality, too. This issue regards the working hours of the employee. Pay observations—including for hourly pay—can be taken simply over the head count of employees or alternatively over the count of hours worked, that is over employees weighted by their working hours. The latter boils down to full-time equivalent wage levels and lends part-time employees a lesser weight in determining the average and the quantiles. Evidently, such weighting reflects more closely the economics of the labor market and less the receiving side of labor's personal incomes, which affect their significance for household welfare and spending; both sides deserve consideration, and attention should not focus exclusively on one or the other.

Finally, there is yet another timing issue on the employee side: wages can concern all who are in work during a year or they may be restricted to those who work the full year, or alternatively all workers may be considered in terms of full-year equivalents. Covering all includes the people who enter or leave employment (or both) in the course of the year; in the full-year option they are left out, in the annualized full-year equivalent approach they will be weighted also by the part of the year in which they work. The share of part-year workers naturally differs between social groups, but it can differ also over countries and over time, because of the business cycle or because of a different or changing role of temporary jobs. New entrants in particular may have low wages and significantly affect inequality at the left-hand tail of the distribution. Finally, the part of the year they actually cover—say 3 months instead of four—will affect their earnings considerably and may have a significant effect at the margin on annual earnings inequality.[36]

To conclude, we do not think there is one best definition of the wage or earnings variable—it depends on the purpose of the analysis. We do think that definition and

[36] The time basis of the wage variable should be a matter of concern as it may cause major differences in the level of wage inequality. On an annual basis inequality may be five to six times larger than on an hourly basis (Karoly, 1993, Appendix B2B); the annual dispersion of hours worked explains the difference. Even on the much-used weekly basis there is a clear dispersion in the hours of work (Karoly, ibidem). Second, the dispersion of hours within categories in combination with their weight in the total will affect outcomes, cross section, and over time. For example, men's hours' dispersion seems much more compressed than women's and their compositional weights have developed strongly; in other words, full-time full-year working men are becoming steadily less representative of the wage distribution as a whole.

purpose should be explicit and mutually consistent and that shortcuts adopted for reasons of data covenience should be scrutinized for their hidden properties and potential effects on the outcomes.

18.3.1.3 Composition and Samples of the Population

Another issue worth mentioning is the part of the population that is covered by the analysis. A *pars pro toto* approach, that views a part of the population as representative of the whole, is particularly dangerous in inequality analysis. Subsets of the population may occupy very different positions in the overall distribution, and inequality may differ significantly between them. Their inclusion or exclusion can exert large marginal effects on the level of inequality even when they are relatively small compared to the whole population. Selection along dimensions such as gender, age, education, or experience on the side of the person, or industry, occupation, the nature of the employment contract and its protection, and the (part-time) working hours on the side of the job can greatly affect the aggregate outcome. The issue may seem obvious although it frequently is a source of error, confusion, or even distortion. For example, contributions may focus on men, on people working full time, on the working-age population, or on positive incomes only, as if assuming that all the rest of the population makes no difference to the general outcome nor to that of the selected group. Imagine that women increasingly occupy low-paid jobs while men are ousted and leave employment; both groups could potentially see their wage inequality fall, though overall it might actually increase. Another realistic example is from Krueger and Perri (2006), who draw conclusions about household consumption inequality for the United States as a whole on a (laudably specified) sample that leaves out non-working-age households, those without an income from labor, and rural households—which are groups that may substantially affect inequality at the margin. Finally, even if all the population is covered all the time, compositional shifts across categories may be highly relevant to the evolution of inequality and will need proper scrutiny. Vice versa, aggregate stability of inequality can go together with changes in inequality within many distinct categories; in the extreme case, even all categories could face inequality change in the same direction (together with shifts in their positions relative to each other). Finally, it is important to add the observation—found in the overview of the literature below—that the distinction of between-inequality and within-inequality (the residuals, after all) depends on the variables chosen as the basis for the decomposition. That choice will likely be inspired by what are considered to be the stylized facts; as a consequence, insufficient attention may be paid to the implications of large residuals and these may actually obtain an importance of their own as is underlined by some of the literature that we will be discussing.

18.3.1.4 Data Sources/Statistical Observation

Individual wages seem more cumbersome to observe statistically than household incomes. For incomes, the collection of taxes provides a strong and universal incentive for gathering administrative data. Such data usually combine considerable precision regarding the core variables with clear limitations for other variables such as personal characteristics, for example, educational attainment is of no direct importance to the tax authorities. This motivation may be less compelling for a comprehensive collection of wage data. Administrative data may be gathered for registering individual Social Security entitlements, but their nature and coverage will depend on the idiosyncracies of the entitlement rules; for example, the sampling may be restricted to those who can qualify for the entitlements in question (e.g., after a probation period, working a minimum of hours, excluding overtime earnings), or focus on their work histories and not their actual earnings, or cover their earnings up to a relevant threshold only. This may hamper their use particularly in international comparisons.

Dedicated surveys, by contrast, require a special effort and consequently are subject from the start to cost–benefit trade-offs, which will affect the range of variables, the population samples, and the time periods covered. This explains why surveys may concentrate on information that is easier to collect, and also that significant international differences occur in the availiability of data and in their coverage. As a result, one can understand the long-time focus in OECD data and American data, along with analyses of full-time workers[37]: collecting hours information on top of earnings information to enable determining hourly wages, or information on workers who have left during the year on top of those permanently employed or present at the time of the survey, is simply more demanding and costly. This may be the case particularly if the information is gathered from employers. Note, though, that ICT developments are greatly facilitating the transfer of firm data to statistical offices. Employers will, by their own interest, dispose of the most accurate information about pay. By contrast, if the information is gathered from households, the information on wages will be less precise, as respondents may not know the details of wage components or taxation and contributions, or respondents may actually be less well-informed than other members of the household. Equally, the information about hours of work may differ between employer sources and household sources, as the former will focus on legally formalized working time whereas the interest of the latter will be in the actual hours that a job involves, possibly including the necessary travel times. Interestingly, a concentration on full-time full-year workers may make little sense in a household survey as it will add to the costs. At the same time, employers will be less well informed about workers' personal characteristics such as educational attainment or the worker's household situation, and the availability and quality of that information from a household survey may be superior compared to employer surveys. Another advantage of administrative tax data can be their more comprehensive time coverage—tax is paid over the full year—whereas household surveys may have important

[37] Stretching all the way from the 1980s to Acemoglu and Autor (2011) included. Heathcote et al. (2010, 24) point out the inadequacy of this focus.

limitations, such as what time of the year the survey questions are asked—do the questions relate to the preceding year or the current one? Adding the dimension of hours to that of earnings can only complicate this.[38] Finally, administrative data will normally cover very large shares of the population and ascertain that all essential questions are answered, whereas other surveys can cover only much smaller samples and suffer from considerable nonresponse to questions,[39] generating less-accurate results also as a consequence of that. Nonresponse will be more important for the current focus on wages at the very top; unsurprisingly, tax data play a large role here though the top-coding of responses may still affect the availability of data, but that is no different for wages than for incomes. As administrative data will be available anyway, increasingly, the statistical offices are trying to use these instead of asking fresh questions to households or firms, and use those data for imputations in other surveys, blurring the distinction between the two types of information as a result. Naturally, both administrative and survey data are subject to changes over time. The tax system or Social Security rules may change and ask for new variables or drop existing ones. A survey may be adapted also because of costs, or simply because a new survey is started without paying due attention to the continuity with its precedessors.[40]

Having said this, the main data source in the literature is first and foremost the American Current Population Survey CPS. It is a household survey, started in the 1940s and providing tabulated data from then, that has made microdata available for research since the early 1960s (the more adequate CPS ORG—outgoing rotation groups—data being available since 1979 only). CPS comes in different "tastes": the March CPS or the May and/or ORG CPS, and one needs to be careful which one to use, partly depending on the purpose of its use. The March CPS is not good for hourly wages, whereas the CPS ORG does a better job here and also has a much larger sample size than the May CPS, which, in addition, may be seasonally affected while the ORG CPS data cover the full (preceding) year.[41] However, the practice of top-coding of labor incomes may reduce the usefulness of this source of data for studying earnings inequality.[42] Several other American data sources are sometimes used, such as the PSID (which we will use below

[38] Below we are forced to combine from EU-SILC survey-time working hours with preceding year information; the American PSID is subject to similar problems.

[39] Up to one-third of CPS wage observations may be imputed by the surveyors (John Schmitt at CEPR Washington, DC—personal conversation).

[40] The break between ECHP and EU-SILC is a case in point, but over its long duration the American CPS also shows several important changes.

[41] See also Lemieux (2008) for a detailed discussion.

[42] "For example, in the March CPS, reported wages and salaries were until recently top-coded at $150,000 a year, which is barely above the 95th percentile of the distribution of earnings in the tax data of Piketty and Saez ($125,471 in 2004). One well-known data set for which top-coding is not an issue is the Panel Study of Income Dynamics (PSID), which is unfortunately not ideal either for studying top-end inequality because of smaller sample sizes" (Lemieux, 2008, p. 32).

for better mimicking European SILC data) and the census, and also employer surveys such as the Employment Cost Index microdata (Pierce, 2001, 2010).

Second, on the EU side, increasingly two consecutive EU-wide (panel) surveys provide microdata for research: the European Community Household Survey (ECHP) and the Statistics on Income and Living Conditions EU-SILC. The ECHP covers the EU15 only, with the exception of the first years of Austria, Finland, and Sweden, who joined the EU in 1995. The survey performed eight annual waves in the years 1994–2001, generating annual data for the years 1993–2000. Sample sizes and degrees of panel attrition diverge substantially across countries depending on the value attached to the survey in the country.[43] The ECHP was discontinued and has been replaced with EU-SILC, which is still in force today. SILC has annual waves starting in 2003/2004 and extending to 2012 at the time of writing—again relaying full-year data for the preceding years (in most countries). SILC's country coverage follows the extension of EU membership and attains full coverage of EU-27 together with Iceland, Norway, Switzerland, and Turkey in 2007.[44] There are a host of small differences between countries in sampling, definitions, and the like, and these also change over the years. Importantly, the *gross* wage variable has been available for all countries since the wave of 2011, although up to then some countries provided net wages only (France, Italy, Switzerland).[45]

Another easily accessible and often-used international data set is the OECD's earnings database, which provides tabulated data. It has been built since the mid-1990s and now covers 34 countries,[46] albeit with rather uneven time coverage. Only seven countries go

[43] Particularly, the educational variable suffers from different national interpretations of the common data-gathering conventions in the course of the waves. In France and the Netherlands almost all responderts are misclassified at the lowest level of education from 1997 onward. At the same time the United Kingdom drastically alters its classification of educational attainment with a strong upward effect among the population as a result.

[44] Brandolini et al. (2011) consider SILC data in detail, and also attempt to aggregate over European countries. See epp.eurostat.ec.europa.eu/portal/page/portal/microdata/documents/SILC_IMPLEMENTATION_headezr.pdf

[45] The most advanced experiment in income and wealth data harmonization is known under the old name of Luxemburg Income Study-LIS (http://www.lisdatacenter.org/). LIS is home to two databases, the *Luxembourg Income Study Database* and the *Luxembourg Wealth Study Database*. The income data set contains information for 46 countries, in some cases going back to the 1970s. A parallel project was started at Cornell University, known as Cross-National Equivalent File-CNEF, 1970–2009, in collaboration with other research partners (see http://www.human.cornell.edu/pam/research/centers-programs/german-panel/cnef.cfm). The Cross-National Equivalent File 1970–2009 contains equivalently defined variables for the British Household Panel Study (BHPS), the Household Income and Labour Dynamics in Australia (HILDA), the Korea Labour and Income Panel Study (KLIPS) (new this year), the American PSID, the Russian Longitudinal Monitoring Survey (RLMS-HSE) (new this year), the Swiss Household Panel (SHP), the Canadian Survey of Labour and Income Dynamics (SLID), and the German Socio-Economic Panel (SOEP).

[46] The usual suspects from America, Asia, and Europe together with Chile, Iceland, Israel, and Turkey.

back in time before 1990, and complete coverage is very recent (2010). In most cases the data are provided by the national statistical offices, although in a few cases they are derived by the OECD from other surveys or provided by national experts. However, definitions and samples vary widely between countries, covering the entire set of possible differences that we have just discussed, ranging from all all individual employees to full-time, full-time full-year employees, and full-time equivalent employees, from hourly wages to weekly, monthly, and full-time equivalent annual earnings, and from gross to net after taxes and contributions. The latest version of the full database contains 90 different series endorsing 33 different definitions. It commonly details the outcomes also for the two genders. For the website version of the database, the OECD has chosen to present only one series per country, 33 in total. This reduces diversity to nine different definitions; the mode (20 series) concerns full-time employees' weekly or monthly gross earnings (which may be deemed reasonably comparable[47]) but only 11 of those go back in time before the year 2000. All definitory properties are admirably documented in the database and offer the user the opportunity to consider the differences and their potential effects. Nevertheless, the database is clearly not immune to the problems of secondary data sets that Atkinson and Brandolini (2001) have stressed for incomes, but which are equally important for wages and earnings.[48]

Finally, Atkinson (2008) provides the results of an in-depth study of the earnings distribution in 20 countries, inspired by the work of Harold Lydall (1968). He advocates a long-run picture on a year-by-year basis, showing that "drawing on isolated years . . . can be misleading." For each country an extensive appendix documents the available data sources and the properties of the data and the presents the evolution at various percentiles of the distribution, ranged separately for the lower and the upper half of the distribution. The series end in 2004 and stretch back in time to well before those of the OECD database. For 15 countries they start before 1960 and cover most of the postwar period and some of those (Canada, France, Germany, United States) go back to before the war.[49] This long time span helps to realize the particular nature of the more recent developments that are the subject of the debates considered in this chapter. Roughly speaking, strong declines in inequality over previous decades preceded the

[47] Often a week is taken as 4/13th of a month, or vice versa.

[48] They discuss other attempts of international data gathering apart from the OECD's, consider some of their use in the literature, and list the factors that influence what they call "a bewildering variety" of inequality outcomes. It is highly important to consider the variation and its implications when using the data for international comparisons of levels as well as evolution. Atkinson and Brandolini "caution strongly against mechanical use of such data sets." They also mention that country fixed effects may not provide a remedy and that even when data are uniformly defined the precise definition may have an effect on the conclusions that can be drawn. Atkinson (2008) extensively discusses similar issues with a focus on earnings, and adds important detail by wage definitions and time periods for 20 countries out of the OECD's 28.

[49] Atkinson and Morelli (2012) update the P90:P50 ratio to more recent years for most of these countries and add a few other countries.

increase on which the literature started to focus in the 1980s. Preferably, the analysis should be able to also explain the declines.

To conclude it seems natural that contributors to the literature are requested to specify their definitions, samples—including censoring or top-coding—as well as sources. Given the long history of using the CPS this is increasingly becoming standard practice in American contributions, but it certainly needs endorsement in international comparisons. Equally important, but not frequently practised, it seems highly advisable to consider the possible implications that data limitations and data choices made may have for the conclusions that are drawn.

18.3.2 Cross-Country Levels and Evolution of Wage Inequality

We now turn to the stylized facts of earnings inequality as we derive them from the literature. This is done in two steps. We start with the United States, which is the country having the best information and where the debate and the analysis of earnings inequality have developed most strongly, enabling us to spell out most of the issues at stake. We contemplate the variation in outcomes between different measures of inequality where feasible, between different definitions of the wage variable where necessary, and between different data sources where reasonably available. In addition to discussing aggregate outcomes, we take a look at some breakdowns—both of the earnings distribution itself and by segments of the (employee) population. Next to the United States, we continue with a consideration of various other countries aimed at comparing the inequality trends but also at identifying gaps in the available data that hamper comparability. In Section 18.3.3 we provide some new empirical evidence from a cross-section comparison of the EU countries and the United States for the most recent year available, based on EU-SILC and PSID, which we will use for our empirical approach in Section 18.5. We end with summary conclusions regarding the stylized facts in international comparison.

18.3.2.1 U.S. Earnings Inequality

Much of the American literature focuses on men or at least distinguishes between the sexes, treating them separately and seldom putting them together in the overarching distribution. This contrasts with other countries and seems a paradox as US female employment started growing earlier than elsewhere and also grew more fiercely in the sense of being predominantly full time and extending high up the overall earnings distribution (Salverda et al., 2001). It may be explained from the early start of the inequality debate in the United States at a time that data did not really allow putting them together. This split risks ignoring the genders' mutual interaction in labor supply and demand and overlooking also the contribution of the within-country doubling of the labor force between the late 1960 and mid-2000s, which has remained in the shadow of the worldwide Great Doubling, a term famously coined by Freeman (2006). For this reason and for the sake of international comparability, and also because it allows covering the recent years since the mid-2000s, we start with a quick look at the aggregate level of all employees irrespective of gender. That comprehensive picture is provided by Figure 18.11. Panel A indicates the overall percentile ratio, P90:P10, from two different sources, the EPI's *State of Working America* and the OECD's Gross Earnings Database. EPI covers hourly wages of all employees, presumably based on head count individuals and not full-time equivalents; the OECD data, by contrast, concern weekly earnings of full-time employees and therefore miss out on part-time employment.

Starting at exactly the same level in 1973, EPI shows a much stronger increase in the ratio between 1979 and 1988 than OECD, directly followed by a decline while the OECD series remains unchanged. At the end the inequality level according to EPI is well below the OECD's.[50] The conceptual difference between the two is important as it is found throughout the literature. Lemieux (2010) as well as Heathcote et al. (2010) provide state-of-the-art overviews of developments for many aspects of American earnings inequality from around 1970 to the mid-2000s, entirely based on hourly wages (but always split by gender).[51] Other important contributions (e.g., Acemoglu and Autor, 2011), by contrast, draw to an important extent on full-time weekly or full-time full-year workers (equally split by gender). Autor et al. (2008, Figures 2 and 3) draw a useful comparison between hourly and full-time full-year earnings inequality trends.

Panel (b) pictures the percentile ratios of the common split between the upper and the lower half of the distribution from the same two sources. It suggests that the difference between the sources and the definitions concentrates in the bottom half; in the upper half the two series are almost identical, which is understandable as in this case virtually all employees will be working full time. The divergence between the two halves is an important observation to retain. The panel also suggests, in accordance with much of the literature using the gender breakdown, that developments since the early 1990s have been different from before, because, on the one hand, lower-half inequality hardly changes in contrast to the preceding period, but upper-half inequality keeps on growing relentlessly, with ends far exceeding bottom-half inequality. With the EPI data, the divergence starts in 1992, with the OECD in 1995.

Finally, panel (c) adds a rather different way of presenting the evolution of inequality: the cumulative changes in real wage levels for each of the 100 percentiles over different time periods, using the work by Pierce (2010). This has become a convenient way of presenting the data in the polarization debate that we will report on later. The discontinuous periodization highlights apparent differences but may suffer from a certain arbitrariness at the same time. With its detail, this type of presentation seems to implicitly

[50] Our aim is not to seek an explanation; a possible one may reside in the variation in full-time hours across individuals. See Autor et al. (2008, Figure 3) for outcomes similar to panels (a) and (b).

[51] The two seem rather different at first sight; e.g., Lemieux finds much lower levels of the variance for both males and females; however, from 1979 to 2005 the trends are largely identical.

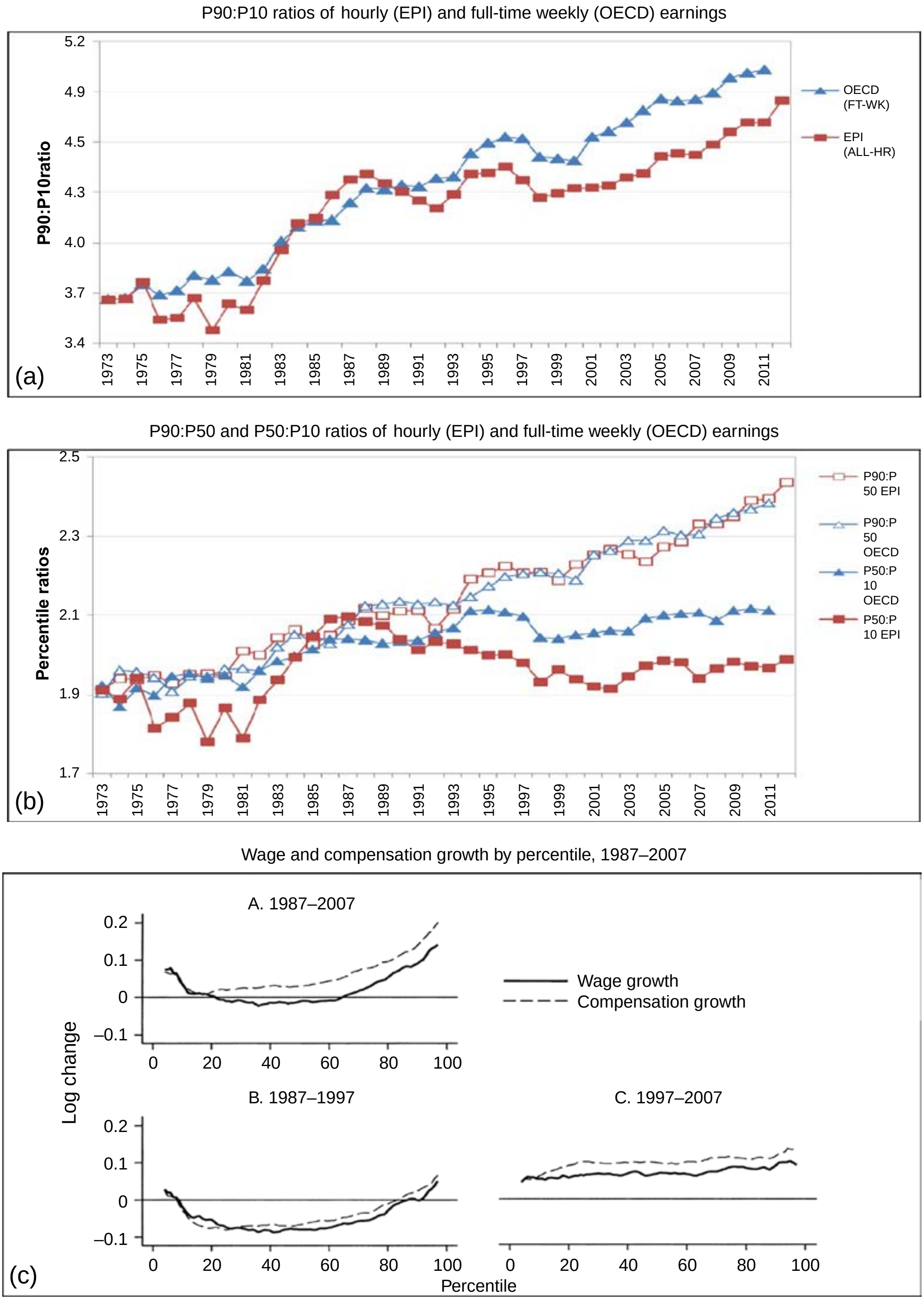


Figure 18.11 Inequality of individual earnings, United States, 1973–2012. (a) P90:P10 ratios of hourly (EPI) and full-time weekly (OECD) earnings. (b) P90:P50 and P50:P10 ratios of hourly (EPI) and full-time weekly (OECD) earnings. (c) Wage and compensation growth by percentile, 1987–2007. Sources*: OECD Gross earnings: Decile ratios (26 October 2013), Economic Policy Institute EPI,* State of Working America 2012*, Washington: Real wage deciles, all workers, 2012 dollars (based on data from the CPS), and* Pierce (2010*, Figure 2.5) (based on Employment Cost Index data).*

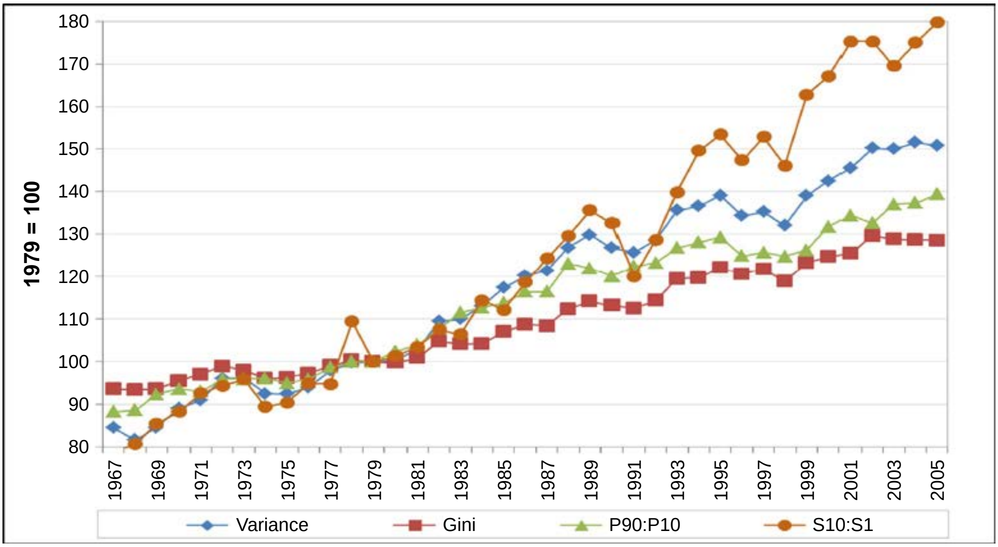


Figure 18.12 Four measures of hourly wage inequality: United States, Men only, 1967–2005, 1979 = 100. *Reading note*: S10:S1 is the ratio between average hourly wages in the top decile to the bottom decile. *Explanatory note*: Figures cover individuals aged 25–60 who work at least 260 h per year, with wages at least half of the legal federal minimum wage. Source*: Heathcote et al. (2010, Figure 4) and S10:S1 derived from Figure 7 (based on March CPS).*

criticize the use of more aggregated measures such as the Gini coefficient or the overall percentile ratio. The panel shows a much flatter pattern of changes over the 1990s than over the 1980s, when strong declines in real wages occur for most percentiles between the tails of the distribution. Nevertheless, real wage growth mostly increases with the wage level. Interestingly, the panel elaborates also on total compensation (dashed lines), which includes employer contributions on top of wages. This is a unique feature that will be mentioned only here. We may conclude from it that the comprehensive concept of earnings does not change the general patterns for the 1980s and 1990s though it reinforces inequality levels somewhat during both periods.[52]

Figure 18.12 draws a comparison (for men only) of the intuitive overall percentile ratio with the often-used aggregate measures of log wage variation and the Gini coefficient and also with the ratio of average wages in the top and bottom decile (S10:S1). All measures show much higher levels now than in the 1970s. However, the variance grows substantially more strongly than the Gini coefficient, while the percentile ratio fluctuates

[52] Congressional Budget Office (2012, Table 7) suggests (for annual earnings) that the relative top-up of cash wages and salaries with contributions to deferred compensation and employer contributions to health insurance and payroll taxes has grown over the quintiles of employee incomes.

between them, the S10:S1 ratio runs away from the rest after 1993.[53] The divergence between the S10:S1 ratio and the common P90:P10 ratio implies that the rapid rise has to do with the within distribution of the two tails, which none of the other three measures seem to be able to capture adequately. The top-incomes literature has already shown its importance at the top end, but the dispersion within the bottom decile merits equal attention.[54] Apparently, the strength of the increase in the dispersion depends on the measure chosen and also their periodic ups and downs do not fully coincide.

The top and bottom half split of Figure 18.11 has provided a first breakdown of the aggregate by focusing on parts of the distribution. The incidence of low pay or high pay, and the size of the remaining middle are indicators of the same sort. The former was already shown in Figure 18.10. It moves up from 23% in 1975 to 25% in the mid-1980s and has been rather stable at about that level since. Over the same period the share of those high-paid, defined as earnings exceeding 1.5 times the median hourly wage, increases from 21% to 25% in the mid-1980s and further to 28% (not shown). As a result the remainder in the middle shows a considerable fall before the mid-1980s (55–50%) and another slighter fall over the current crisis (49–47%). A narrower definition of high pay following the top-incomes literature is pursued by Lemieux (2010), who endorses a simple repair for the top-coding of earnings in the CPS[55] and presents percentiles distributions similar to those of Pierce above, which we reproduce in Figure 18.13. Starting in 1974 the period covered is significantly longer but still split into two parts, now on both sides of the year 1980. Separate distributions are given for men and women. Again developments are more positive and spread more evenly over most of the distribution during the second period after 1989 than before. The longer period covered up to 1989 shows a more skewed picture than Pierce's. Particularly, real wage change in the bottom 20% seems more negative now for men, although an increase in the lowest percentiles for women may help explain the surprisingly upward move found by Pierce. At the same time, it is clear that among men the high part of the distribution has run away from the rest with a steep gradient within the top decile. The top percentile ratios seem to support this (not shown). They are almost identical and trend upward together until the end of the 1990s when female inequality starts to lag behind. The bottom-half ratios run largely parallel to each other with the one for females indicating a substantially lower level of inequality. The more positive development of wages for women seem suggestive of a declining gender gap. This is borne out clearly by Heathcote et al. (2010, Figure 5) who, after a slight increase of the gap from 1967 to 1978, find a continuous decline after

53 The evolution of the S10:S1 ratio seems to imply that the variance is a plausible measure to use here in spite of its sensitivity to outliers in the distribution.

54 Since the mid-1980s the incidence of the minimum wage ranges entirely within this decile.

55 Checking against the Pareto parameter-based approach of Piketty and Saez (2003) he concludes that results are the same.

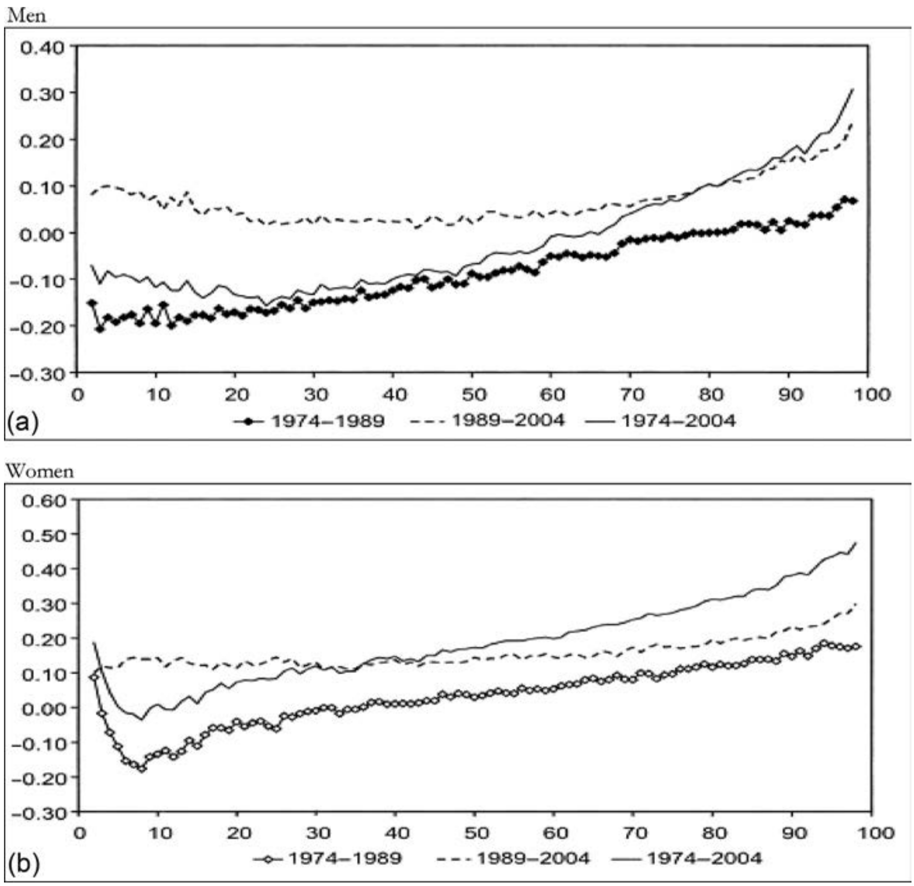


Figure 18.13 Percentage change in real hourly wages, by gender and percentiles, United States, 1974–2004. (a) Men and (b) women. Source*: Lemieux (2010*, *Figure 1.7).*

that year, sharply up to the mid-1990s and more modest since then. The current gap (30%) is much smaller than before but certainly not negligible.

Next to gender, educational attainment is the most important dimension for breaking down inequality. Its role has been a bone of contention in the literature from the start, as we will see later. Here Lemieux (2010) presents differentials for various levels of attainment relative to high school graduates (Figure 18.14). They appear to be mostly flat with slight declines at the lower levels but with the clear exception of the highest two levels, particularly the highest. These start growing away upward particularly over the 1980s and more modestly since. For men the top–bottom gap almost doubles. At the end of the period the differentials seem almost identical between the two sexes. Heathcote et al. (2010, Figure 5) present a college wage premium defined as the ratio between the average

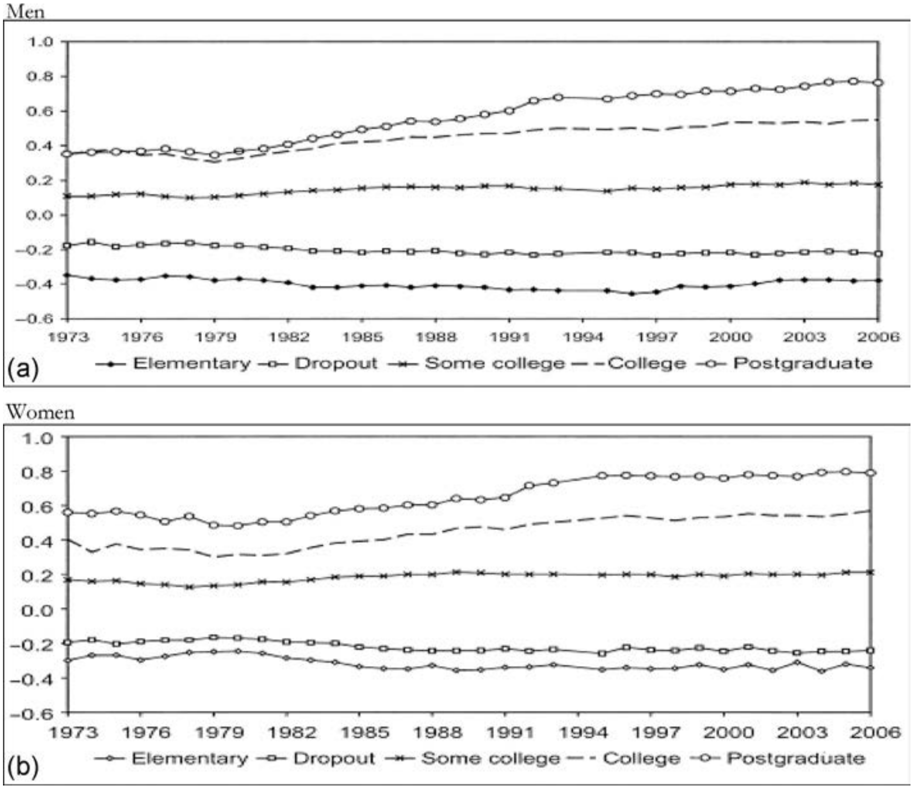


Figure 18.14 Educational differentials relative to the high-school level, by gender, United States, 1973–2006. (a) Men and (b) women. *Explanatory note*: Using a decomposition based solely on education and experience. Source*:* Lemieux (2010)*, Figure 1.3.*

hourly wage of workers with at least 16 years of schooling, and the average wage of workers with fewer than 16 years of schooling. The premium increases significantly though more for men (52–92%) than for women (58–69%).

We stop here and refer for further detail of other dimensions of the earnings dispersion, such as experience or nationality/country of birth, to the literature itself. Before we continue we will stress again the important role of residuals. These outcomes for gender and education rest on simple decompositions, and most of the action appears to reside in the residuals, which develop largely in parallel to the growth in overall "raw" inequality (Heathcote et al., 2010, Figure 5). The implication is that other factors of influence need to be incorporated in the analysis and/or that idiosyncrasies, which may be immune to further analysis, play a nonnegligible role. Lemieux (2010, Figure 1.8) finds, interestingly, that the importance of residuals grows with the level of earnings, especially over the period 1974–1989.

18.3.2.2 Earnings Inequality in Other Countries

We now turn to inequality trends in other countries. The *Review of Economic Dynamics* (RED) (2010) is a special issue dedicated to cross-sectional facts regarding elements of economic inequality; it provides the most precise cross-country comparison of the earnings dispersion and several of its important facets, using as uniform a template of data treatment and presentation as possible.[56] Unfortunately, it has several drawbacks. The limited number of countries of relevance here is only seven, and for our comparison it is even further reduced as Italy and Spain focus on earnings net after taxes, which are unsuited for a comparison to gross earnings, and relevant data for Sweden end in 1992 when the country's financial crisis had just started. That leaves us with the United Kingdom, Canada, and Germany. We turn to these results first, and after that we turn to the OECD's earnings database to see what we can learn for other countries.

Figure 18.15 presents three measures of individual hourly earnings dispersion, for men and women together, as found in the RED contributions: the variance of log wages, Gini coefficient, and overall percentile ratio. All indicators for the three countries tend to rise over time. The British variance increases very rapidly up to a level 60% above the start of 1978, which well exceeds the other two measures (and also the variance in the United States), and subsequently falls over the 2000s. The other two measures for the United Kingdom also show a decline over that last period. This contrasts with the OECD's percentile ratio (not shown), which (covering full-time weekly earnings) is at a somewhat lower level but continues rising until 2006 and remains unchanged until 2011. For Canada the rise is also considerable, +20–40% depending on the measure, and continues until the end of the period. Mutual differences between the measures are smaller. The OECD's percentile ratio (not shown), again for full-time weekly earnings and available after the mid-1990s only, shows continued growth over the entire 2000s. Finally, in Germany, data are available from 1983, the rise of the three indicators concentrates in the period after unification. The variance shows a clear rise, and it is virtually identical for the percentile ratio—in contrast with the other two countries. Taken over the same 1983–2004 period, their growth is stronger than in the United Kingdom or Canada. The rise of the percentile ratio must rest on the use of hourly earnings as the trend of the OECD's ratio (not shown), which concerns full-time monthly earnings, is largely flat over the 1990s and early 2000s. However, the increase in the Gini coefficient is modest relative to the other indicators as well as the other countries.

[56] Brzezowski et al. for Canada: 1978–2005, Fuchs-Schundeln et al. for Germany: 1984–2005, Japelli and Pistaferri for Italy: 1980–2006, Pijoan-Mas et al. for Spain: 1994–2001, Domeij and Flodén for Sweden: 1978–2004 (effectively 1975–1992 only), Blundell and Etheridge for the United Kingdom: 1977–2005, and Heathcote et al. for the United States: 1967–2005 (but always split by gender). Unfortunately, Spanish data are net after taxes and will be left out here. Krueger et al. provide a summary overview in the Introduction. Individual earnings dispersion is addressed as part of the study of the household distribution of earnings (mentioned above, see, e.g., Figure 18.4).

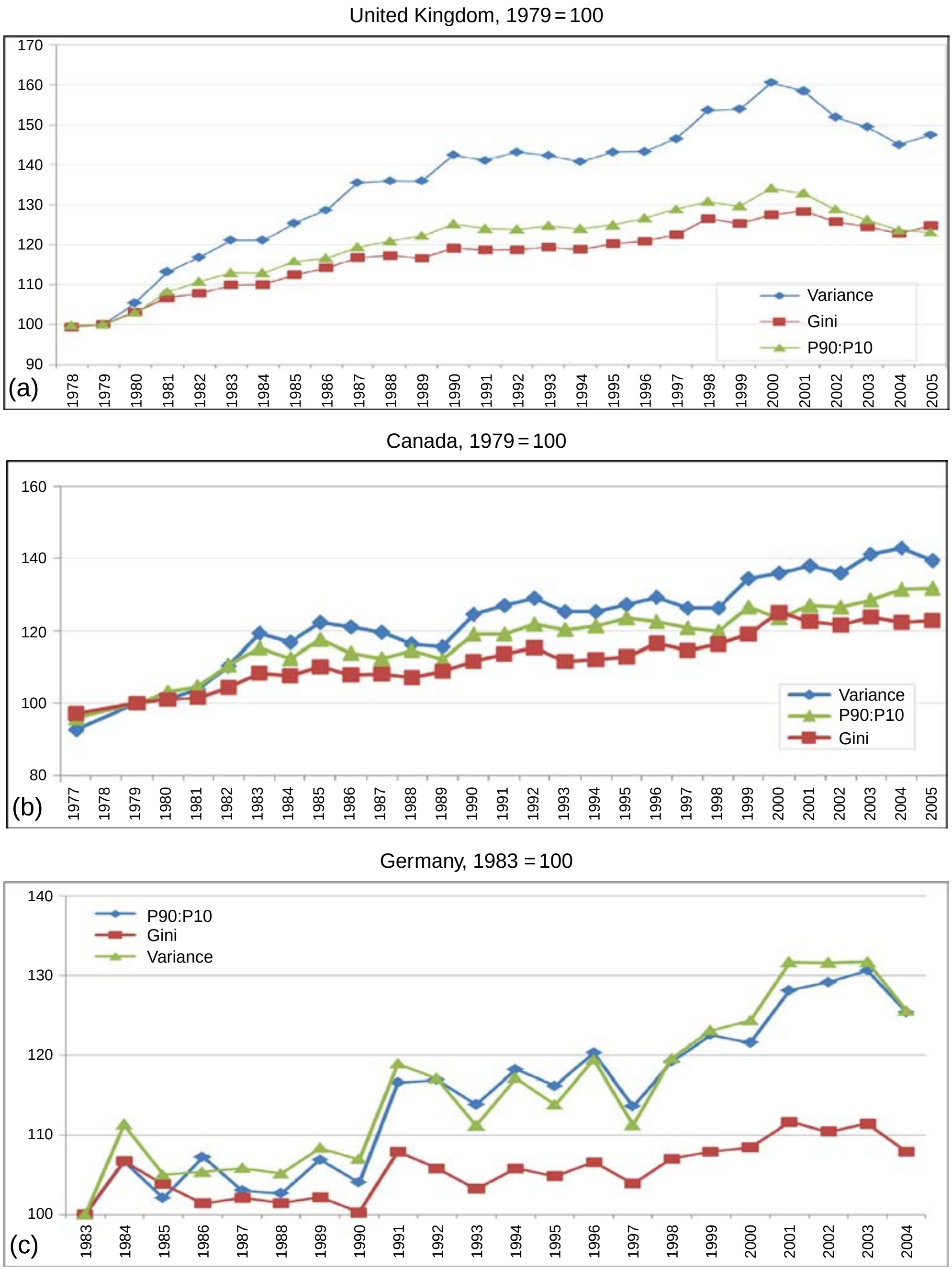


Figure 18.15 Dispersion of hourly wages, United Kingdom, Canada and Germany, late-1970s to mid-2000s. (a) United Kingdom, 1979 = 100. (b) Canada, 1979 = 100. (c) Germany, 1983 = 100. *Explanatory note*: Figures cover individuals aged 25–60. Derived from the data underlying Figures 3.1 (UK), 4 (Canada) and 6 (Germany) as posted on the RED website. Data sets concern for Canada the *Survey of Consumer Finances* SCF 1977–1997 and *Survey of Labor and Income Dynamics SLID* 1996–2005, for Germany the German Socioeconomic Panel GOEP, and for the UK the Family Expenditure Survey FES and Labor Force Survey LFS. Sources*: Brzozowski et al. (2010), Fuchs-Schündeln et al. (2010), and Blundell and Etheridge (2010)*.

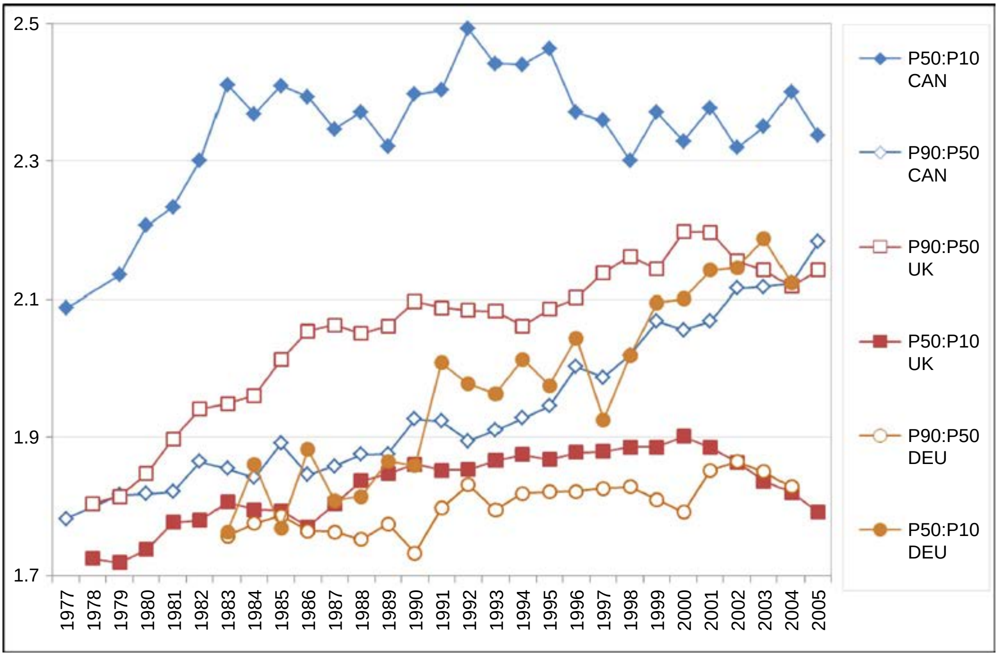


Figure 18.16 Lower and upper half inequalities in Canada, Germany, and United Kingdom, 1977–2005. Sources*: Blundell and Etheridge (2010), Brzozowski et al. (2010), and Fuchs-Schündeln et al. (2010).*

In Figure 18.16 we split the overall percentile ratio into its two contributing halves, P90:P50 and P50:P10. We a find strong divergence between the countries. The high levels and strong rise of the bottom-half ratios in Canada and Germany[57] are strikingly different from the United Kingdom; the German ratio moves up as much as the British but over a considerably shorter period. Canada and the United Kingdom share a decline in recent years though. Upper-half inequality rises very little in Germany and clearly less than in Canada and the United Kingdom. Generally, the pattern of the two British trends is very similar to the United States in Figure 18.11, whereas Canadian trends look surprisingly different. This clearly call for further scrutiny.[58]

Unfortunately, wage changes by percentile, which have come to play an important role in the American debate, are not available for other countries. The RED papers have also looked at the roles of gender and educational attainment. The gender pay gap for Canada is small from the start, comparable to the US gap at the end of the period

[57] For Germany this rise is absent in the OECD's (full-time) data, and therefore also in the overall ratio that was mentioned.

[58] Fortin and Lemieux (2014) conclude to a role of provincial minimum wages and natural-resources growth.

(30%), and it trends downward only very slowly. The German gap declines slightly more steeply and ends below (20%) the American level in the mid-2000s. The British gap, finally, declines somewhat more strongly; it is below the US level in the late 1970s but ends at about the same level. Note again that the decompositions made in the RED papers are based on gender, educational attainment, and experiences only and that in these three counties, as in the United States, residuals are quantitatively important and behind most of the increase in inequality.

With the help of the OECD's earnings database of percentile ratios, the evolution of individual earnings inequality can be described for a number of other countries (see Figure 18.17).[59] As already stated it is a secondary database, and it has to be used with great care. It comprises a wide array of wage definitions and concomitant samples of the employee population. We first select eight countries that focus on gross earnings of full-time workers (be it hourly, weekly, monthly, or annually[60]) and also have a long-run series. Given the diverging incidence of part-time employment, the full-time focus will be more or less representative of the country, but there is nothing we can do about that apart from being aware that in (various European) countries where part-time jobs have become more important and tend to be overrepresented in the low-wage segment of employment, the actual picture of inequality will plausibly be more pessimistic both in cross section and over time than found here for full-time workers only. With the exception of Japan and Finland over the period as a whole and Korea over its first half, overall inequalities in panel (a) seem to be trending upward, albeit to varying degrees and with different timings. Compared to the rest, Hungary and Korea show strong episodic changes, which apparently hang together with deep political change—the end of communism and of dictatorship respectively. The two halves of the distribution are pictured in panels (b) and (c). With the exception of Hungary and Korea, differences in trends seem to be relatively small. Lower-half inequality is usually less than upper-half inequality, and most of the overall rise can be attributed to the upper half. In the stable cases of Japan and Finland, lower-half inequality declines somewhat, but in all other countries it grows at some point in time. Comparable information about gender and educational differentials is not available.

Finally, Figure 18.18 assembles remaining short-run information on gross earnings. This concerns full-time workers with the exception of Denmark (all workers, headcount) and Norway (all workers, full-time equivalents). Countries seem to move into a closer band: higher-inequality countries move down (Portugal, Poland, Greece, Spain) while most of the rest moves upward. Breaking down into halves (not shown) the strong declines in Portugal and Poland are due primarily to the upper half, although lower-half

[59] See Blau and Kahn (2009) for a more detailed international analysis based on this data set.

[60] As a result levels cannot be precisely compared cross-country. Finland, the Netherlands, and Sweden sample full-year workers, which may partly explain their relatively low levels of inequality.

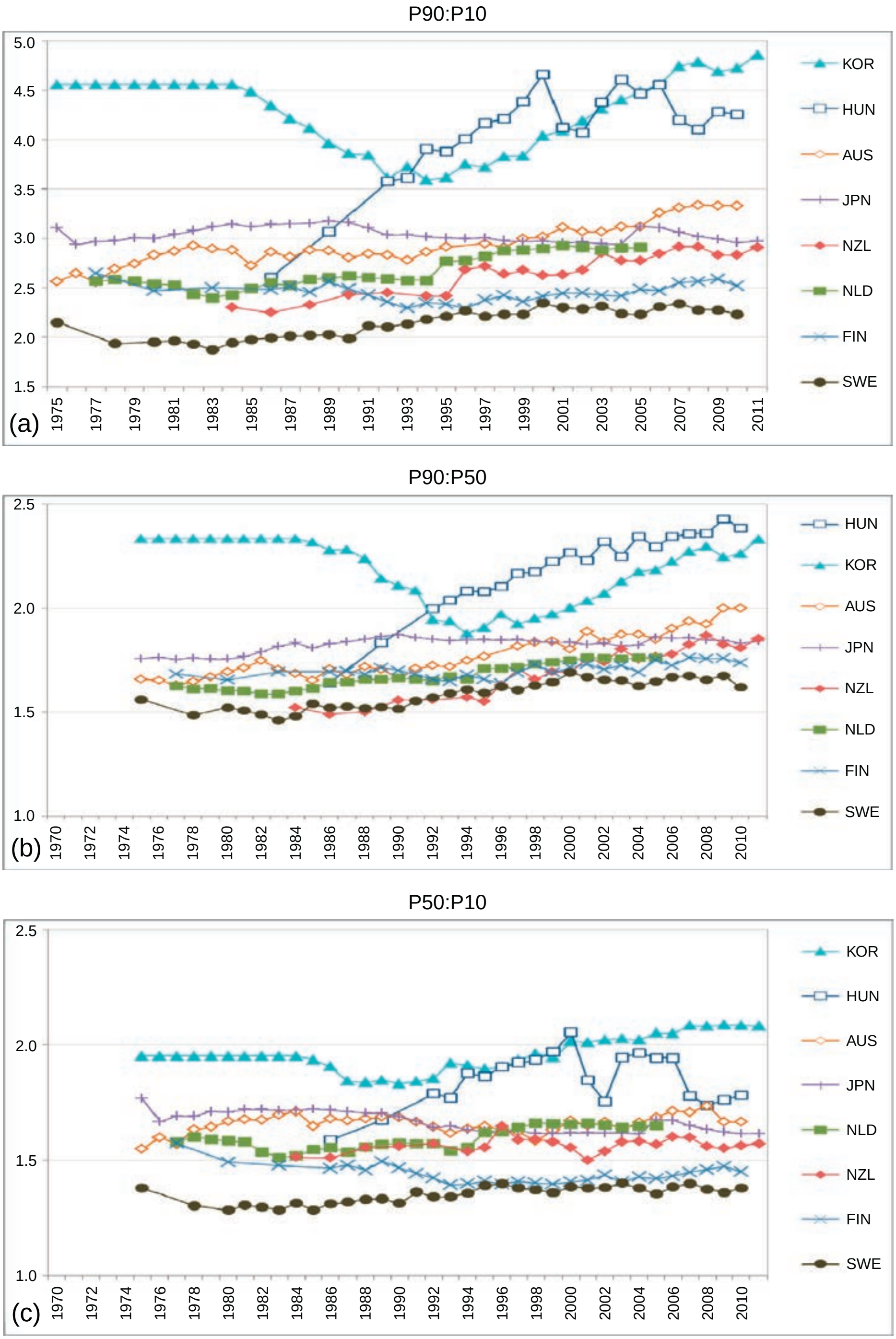


Figure 18.17 Earnings inequality trends in eight OECD countries, 1975–2011. (a) P90:P10. (b) P90:P50. (c) P50:P10. *Explanatory note*: full-time workers only; hourly earnings for NZL, weekly for AUS, monthly for HUN, JPN, ad KOR, full year for FIN, NLD and SWE. Source*: OECD Earnings decile ratios database.*

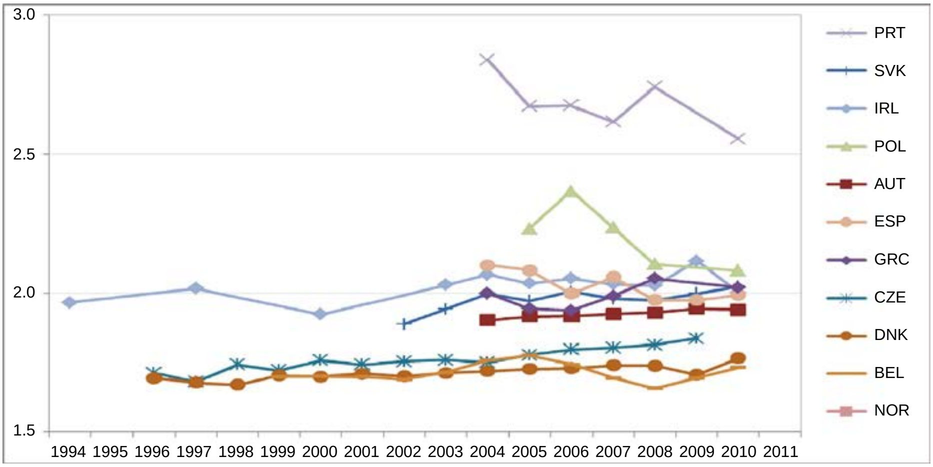


Figure 18.18 Short-run earnings inequality trends in 11 OECD countries, 1994–2011. *Explanatory note*: Gross earnings only; full-time workers except DNL (all, head count) and NOR (all, full-time equivalents). Source*: OECD Earnings decile ratios database.*

inequality also falls. Portugal combines extremely high levels in the upper half with low levels in the bottom half. For the rest, increases and decreases seem split roughly equally between the two halves.

18.3.3 Additional Evidence on Earnings Inequality in European Countries and the United States

From the data that we will be using in Section 18.5 we have derived a cross-section comparison for the most recent year, which provides a useful complement to the above stylized facts. It covers 27 EU countries, Iceland and Norway, as well as the United States. First, we consider a selection of inequality indicators, keeping in mind the household context that was discussed in Section 18.2. In the analysis of income inequality it is common practice to make use of the *Gini concentration index* or, to a lesser extent, of the *mean log deviation* (thanks to its property of decomposability). In earnings inequality analysis, by contrast, the most common indicator is the *standard deviation of log earnings* and/or the *decile or percentile ratio*. In Figure 18.19 we show that different measures provide largely similar country rankings in a cross-country perspective, whereas Table 18.1 provides the correlation indices for the same variables. As known from the literature the first two indices look at the bulk of the distribution, whereas the other two emphasize better what is happening at the tails (Cowell, 2000—see also Heshmati, 2004).[61]

[61] We speculate that the rather lower level of correlation found for the standard deviation may be attributable to top incomes.

Figure 18.19 Alternative inequality measures for full-time employees, EU countries, Iceland, Norway, and United States, 2010. Note and Source: *See Table 18.1*.

Table 18.1 Cross-country correlation indices of various inequality measures, annual earnings of full-time employees, EU countries, Iceland, Norway, and United States, 2010

	Gini index	Mean log deviation	Standard deviation of logs	Percentile ratio 90/10
Gini concentration index	1.000			
Mean log deviation	0.871	1.000		
Standard deviation of logs	0.409	0.770	1.000	
Decile ratio 90/10	0.790	0.855	0.689	1.000

Explanatory note: Full time is defined as working 1000 h per year or more.
Source: Authors' calculations on EU-SILC 2010 and PSID 2011.

Inasmuch as the tails of the distribution may be affected by the increasingly diversified regimes in working hours, we prefer to work with the Gini concentration index, and we provide evidence of various inequality dimensions with the help of this measure. We start with a country overview, as reported in Table 18.2. The first column shows the level of inequality associated with labor earnings, which here include gross earnings from employees and the self-employed together with benefits received by the unemployed.

Table 18.2 Inequality measures for individual annual and hourly earnings and hours, EU countries, Iceland, Norway, and United States, 2010

	Gini of annual labor earnings (including self-employed and unemployment benefits)	Gini of annual gross earnings (hours > 1000) dependent employment	Gini of annual hours worked (positive values)	Gini of hourly gross wages dependent employment	Correlation of hours and hourly wages
Austria	0.376	0.322	0.167	0.325	−0.162
Belgium	0.332	0.257	0.173	0.266	−0.284
Bulgaria	0.422	0.338	0.084	0.318	−0.150
Cyprus	0.392	0.352	0.130	0.350	−0.077
Czech Republic	0.341	0.277	0.095	0.252	−0.124
Denmark	0.278	0.240	0.108	0.228	−0.190
Estonia	0.412	0.342	0.117	0.351	−0.208
Finland	0.361	0.333	0.156	0.340	−0.271
France	0.376	0.328	0.176	0.321	−0.201
Germany	0.420	0.322	0.181	0.307	0.038
Greece	0.485	0.335	0.146	0.338	−0.245
Hungary	0.392	0.351	0.084	0.317	−0.004
Iceland	0.338	0.328	0.172	0.337	−0.196
Ireland	0.448	0.361	0.210	0.374	−0.216
Italy	0.407	0.323	0.135	0.308	−0.230
Latvia	0.494	0.425	0.114	0.401	−0.124
Lithuania	0.494	0.412	0.099	0.403	−0.144
Luxembourg	0.428	0.357	0.166	0.355	−0.145
Malta	0.347	0.276	0.109	0.285	−0.156
Netherlands	0.359	0.292	0.174	0.290	−0.123
Norway	0.329	0.304	0.134	0.287	−0.115
Poland	0.464	0.342	0.122	0.354	−0.178
Portugal	0.453	0.380	0.104	0.374	−0.144
Romania	0.419	0.270	0.046	0.271	0.045
Slovak Republic	0.347	0.273	0.082	0.253	−0.084
Slovenia	0.397	0.337	0.073	0.314	−0.102
Spain	0.444	0.341	0.142	0.313	−0.209
Sweden	0.321	0.282	0.139	0.336	−0.295
United Kingdom	0.466	0.361	0.193	0.371	−0.094
United States	0.570	0.470	0.164	0.603	0.036
Average	0.408	0.332	0.133	0.331	−0.145

Source: See Table 18.1.

Figure 18.20 Inequality in annual labor earnings, EU countries, Iceland, Norway, and United States, 2010. Source*: See Table 18.1*.

The level found here always exceeds the one pictured in the second column for the earnings of full-time employees who comprise a subset of the population considered in the first column. Manifest country-rank reversals occur, plausibly due to a large share of self-employment (as in the case of Greece, Poland, or Romania—see Table 18.A1 in Appendix B) and/or the combination of the unemployment rate and the generosity of the welfare state (as in the case of the Nordic countries—see also Figure 18.20).

Where countries differ more is in the distribution of working hours: because the distribution of hours worked is much less unequal than the distribution of wages (compare second and third columns of Table 18.2), the inequality in hourly wages (computed dividing yearly earnings by worked hours) tends to mimic the inequality in yearly earnings (correlation coefficient is 0.90).[62] This is another important dimension of inequality in the labor market, because given the existing demand for labor inputs, this work can be accomplished by a variable number of individuals, according to existing labor standards

[62] The US exception is accounted for by the fact that hourly wages for European countries are deduced by dividing annual earnings by worked hours, and in the PSID the interviewees are directly asked about their hourly wage. Using the same accounting procedure would reduce the Gini index on hourly wage for the United States to a more reasonable 0.47.

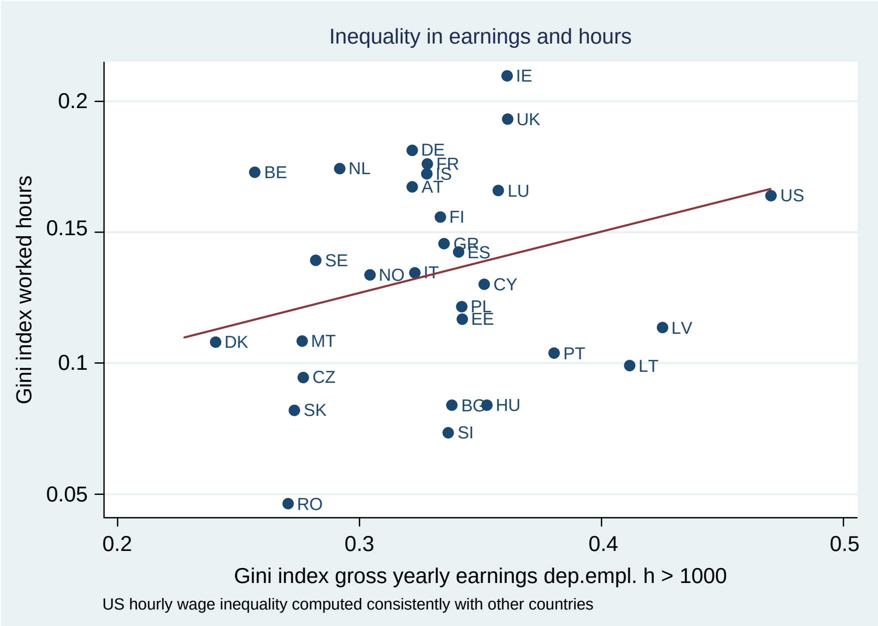


Figure 18.21 Inequality in earnings and hours, EU countries, Iceland, Norway, and United States, 2010. Source*: See Table 18.1.*

and cultural attitudes (regarding female participation, labor sharing within the couples, retirement rules, and so on).

As can be seen from Figure 18.21, the distribution of work may contribute to global earnings inequality, despite being the lowest in formerly planned economies (especially Romania, Czech and Slovak Republics, Hungary, and Slovenia). When a job is characterized by full-time working hours the contribution of working hours to inequality is minimal; by contrast, when flexibilization of the labor market allows for various regimes of working hours (as in Ireland or Great Britain, but consider also the Netherlands, where part-time jobs are widespread), it contributes to the observed inequality in individual annual earnings (which can be partially mitigated by household dynamics, as previously discussed in Section 18.2).[63] However, the picture obtained here by means of aggregate indices is purely impressionistic, as hours and hourly wages tend to be negatively correlated in many countries. As a consequence of the latter, a high inequality in hours

[63] These data on correlation between hours and hourly wages should be taken with caution, because the latter measure is obtained by dividing annual earnings by the former. As a consequence, hours and wages inequality are positively correlated. Thus, any measurement error in the latter generates a measurement error in the opposite direction for the former. However, unless different countries are hit by measurement errors in different (and systematic) ways, cross-country comparisons are still informative of the flexibility in adjustment.

accompanied by a high inequality in hourly wages may produce a low level of earnings inequality; note, however, that this is a possible outcome not a necessary one and that measurement error in hours enhances the risk of spurious correlation.

It is interesting to note that individual workers may react to lower wages by working longer hours—South Korea provides a clear example (Cheon et al., 2013, Figure 2.9). This is consistent with a standard model of labor supply where the income effect dominates the substitution effect. Competing explanations refer to a sort of Veblen's effect: partitioning workers by income layers, if consumption depends on consumption of richer people, an increase in the socioeconomic distance increases the hours worked (Bowles and Park, 2005). The empirical evidence does not contradict this viewpoint (see the final column of Table 18.2, where we have computed the correlation between hours and wages at the individual level). Although the correlation is negative almost everywhere, its intensity varies across countries: in some countries it exceeds 0.20 (notably Belgium, Finland, and Sweden), in other countries it does not differ from zero, suggesting an independent distribution of wages and hours (e.g., United Kingdom and United States as well as Germany—see also Bell and Freeman, 2001). Institutions may be responsible also for this outcome, because employers and workers may have different degrees of freedom in arranging working-hours regime and/or resorting to nonstandard labor contracts. Thus, the two dimensions of inequality (hours and wages) correlate with the same set of institutions, and for this reason in the econometric analysis of Section 18.5 we will allow for this decomposition. But even the correlation between hours and wages itself may be influenced by existing regulations, as it can be considered as evidence of a higher or lower flexibility: the evidence depicted in Figure 18.22 shows that the possibility of adjusting hours when wages are relatively low contributes to reducing earnings inequality.[64] For this reason, in the sequel we will study the correlation between this flexibility measures and LMIs.

The overall picture in terms of earnings inequality is well shown in Figure 18.20: inequality is higher in the so-called liberal market economies (United Kingdom, United States, and Ireland) to which one should add some "transition-to-market" economies (such as Poland, Lithuania, and Latvia) and the Mediterranean countries (Greece, Spain, and Portugal). At the other extreme we find the Nordic countries (except Finland). As Figure 18.23 shows in a clear way, the main determinants of this country ranking derive from the availability of employment opportunities, because countries characterized by high employment rates (including self-employment) are also the less unequal from the point of view of labor earnings. This is partly by construction: because we retain in our sample the entire labor force of the country, whenever the employment rate rises

[64] The United States does represent an outlier, but even after removing this observation, the correlation between the two variables in Figure 18.22 remains positive.

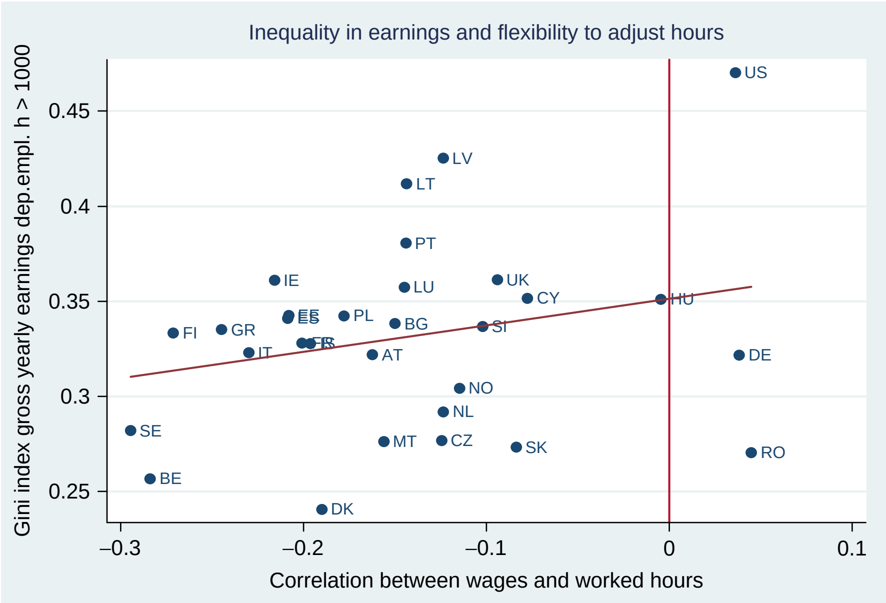


Figure 18.22 Inequality and flexibility for adjustment, EU countries, Iceland, Norway, and United States, 2010. Source: *See Table 18.1*.

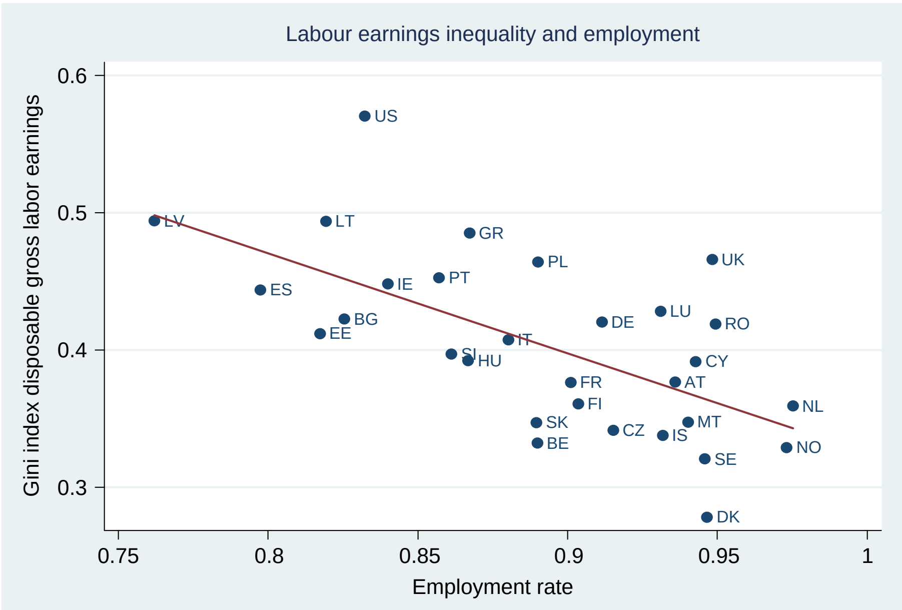


Figure 18.23 Inequality and employment, EU countries, Iceland, Norway, and United States, 2010. Source: *See Table 18.1*.

(and the unemployment rate consequently declines) the measured inequality in earnings declines (see the model proposed in Section 18.5).

18.3.4 Summary Conclusions

A key stylized fact for the United States is that hourly-earnings inequality has increased secularly over the last 40–45 years, not more or less, but more or even more, depending on the inequality measure that is chosen. The rise rests on a virtually continuous increase in inequality in the top half of the distribution; bottom-half inequality grew sharply until the end of the 1980s and after that has remained largely stable. The sharp upward evolution of earnings at the very top, that is reflected by now infamous Top 1% incomes share, makes an important contribution to the continuous rise of that upper half. This is borne out by more detailed changes based on all 100 percentiles that, at the same time, shows some emptying out between the upper and lower tails of the earnings distribution.

A similar divergence between upper-half and lower-half inequality in recent years is found also for the other English-speaking countries (Australia, Canada, Ireland, New Zealand, and the United Kingdom) though the stability of the bottom half may have started somewhat later than at the beginning of the 1990s, or was already there for most of the time. Note, however, that the absolute levels of inequality may differ substantially between these countries, both overall and in the two halves. In some cases bottom-half inequality far exceeds top-half inequality whereas in other countries it is the other way around—naturally the diverging evolution tends to inflate upper-half inequality relative to the bottom half.

The picture is less clear-cut for other countries, ranging from strong increases (Korea and Hungary since the 1990s) to compelling declines (Poland and Portugal in the 2000s; and a small decline for Spain). The comparison is complicated though by international differences in the concept of earnings and therewith in the sampling of the wage-earning population (by the way, also for Australia, Ireland, and New Zealand here above). The sampling often targets full-time employees only, ignoring the part-time ones, who may actually be making important contributions to the level of inequality. Therefore, inequality levels and trends in those countries may be underestimated in comparison. Some countries (Belgium, Finland, and Japan) show flat trends of overall inequality and tend to register declining inequality in the lower half. Most of the rest (Austria, Czech Republic, Denmark, Netherlands, New Zealand, Norway, Slovakia, and Sweden) do show a more modest rise, but a rise nevertheless. In contrast to the English-speaking countries the rise seems to be spread over both halves of the distribution even if the level and the increase are less than in the upper half.

For a few countries the gender pay gap could be consistently compared. This decreased strongly in the United States, more than in the United Kingdom, Canada, or Germany, and from a high level, so that currently, the gaps are of largely the same magnitude. In the US educational differentials between the better-to-best educated and the lower-to-least educated have grown significantly. No internationally comparable stylized facts are available for those differentials though one may assume that they will have grown in many cases albeit to different degrees.

A cross-section comparison of 27 EU countries, Iceland, Norway, and the United States underlines the importance of including the employment chances or, in other words, the distribution of individual hours worked, and not focusing exclusively on the distribution of earnings. The two distributions hang together and do so in different ways, partly depending on LMIs which may make the distribution of hours over individual employees more unequal, e.g., by allowing flexibilization or encouraging part-time hours. Naturally, the effect on household earnings depends on the combination of both hours and wage levels across the members of the household.

18.4. THEORETICAL APPROACHES TO WAGE DISPERSION AND THE ROLE OF INSTITUTIONS

18.4.1 The Wage Inequality Debate 1980–2000 and the Role of LMIs

We start this paragraph with a brief introduction of the evolution of the literature on wage dispersion up to 2000. We go back to the start of the literature on inequality in the 1980s to better understand the current situation and fill a lacuna in existing overviews. This is followed by a more detailed discussion of the study of LMIs in the rest of this introduction. After that we discuss recent contributions in two main directions: supply and demand (4.2) and institutions (4.5); we sum up our findings at the end (4.7).

The literature in which the contemporary discussion on the dispersion of wages is rooted took off seriously in the course of the 1980s, kicked off with a detailed picture of changing male and female inequalities in the United States by Peter Henle and Ryscavage (1980).[65] This literature focused initially on the factual question about whether inequality had increased or not, and it took some time before the factual doubts about that growth dissipated,[66] though factual questions have remained on the agenda throughout. It did not take long before the "why?" question started being asked and answers were sought in many directions—more often than not in different directions at the same time incurring a risk of *ad hocery*. From the start, some of these routes led to what in due course have become known as LMIs. For example, Plotnick (1982) attributed the (slow) increase in the variance of log (male) annual earnings—"earned income"

[65] Henle (1972) already showed an increase in earned income inequality over 1958–1970.

[66] Blackburn and Bloom (1987) compared studies with conflicting outcomes, spelling out important differences in sample choice and definitions, and were about the first to consider both men and women; they concluded that "the time profile of earnings inequality, measured across individual workers, has been quite flat since the late 1960s" (p. 604) combining a decrease in inequality among women with an increase among men.

as it was usually called—between 1958 and 1977 entirely to the differential effects of the level of unionization, the dispersion in weeks worked, the age distribution of workers, and the inequality of education. Dooley and Gottschalk (1984) looked for a demographic explanation (viz. cohort size and the baby boom), and Dooley and Gottschalk (1985) added—on the factual side—the insight that real earnings of men's earnings below a low-pay threshold were lagging behind.[67] Bluestone and Harrison (1982) launched the thesis of deindustrialization and expansion of low-wage employment, which in a later book (Bluestone and Harrison, 1988) turned into the famous "Great U-turn" of growing inequality. This lent political significance to the issue, which was particularly viewed as a polarization negatively affecting the middle class, which may sound familiar to current debates.[68]

Thus, education, institutions, demographics, and the composition of the economy made an appearance in the literature almost from the start. From the early 1990s, international trade and especially the competition with low-wage countries were added as other explanations (e.g., Wood, 1995). However, this is outside the focus on LMIs of the current chapter.[69] Though the interest in demographics may have waned (apart from the fact that gender has become a staple ingredient), the attentiveness to educational differentials, institutions, and the composition of the economy—industries and sectors at the time, occupations and tasks nowadays—has grown into a vast literature over the 1990s. The interest in education focused on the demand for skills in the economy on the one hand and their supply by the labor force on the other hand (e.g., Juhn et al., 1993). This has ushered in the thesis of "skill-biased technological change" (SBTC for short) (Bound and Johnson, 1989, 1992; Levy and Murnane, 1992) as the driving force behind the demand for skills. For some time this became the canonical model for explaining growing wage inequality. Evidently, the composition of the economy is not unrelated to this, if only because the skill structure differs between industries.

Technological change and economic composition together have become the supporting vector for the attention paid in the literature to supply and demand—or "market forces"—as explanatory factors for differences and changes in wage inequality. In their first overview of the literature, Levy and Murnane (1992) recommend future research to "get inside the black box of the firm" (p. 1374) and pursue the hedonic theory of labor demand that views a worker as possessing a fixed package of separate productive abilities and the wage as the sum of payments to them. The worker cannot separate these abilities and sell each to the highest bidder (Mandelbrot, 1962). As a result, the unit price

[67] The threshold was defined as \$3, a good 40% above the Federal minimum wage in 1975.

[68] Harrison et al. (1986) brought the issue to the Congress; presidential candidates George W. Bush and Michael Dukakis discussed the status of the middle class and the nature of job growth in their campaigns of 1988 (Karoly, 1988, 13).

[69] Compare Machin (2008) for an evaluation, finding little support for this explanation.

of a productive ability may vary across sectors (Heckman and Sedlacek, 1985). Basically, Levy and Murnane seem to point in the direction of the route taken more recently by Acemoglu and Autor (2011) with their tasks-based approach. At the same time, it illustrates that technology is difficult to pin down empirically in economic studies and is usually subsumed in the unexplained part of modeling, also in the study of inequality. Although Levy and Murnane focus almost exclusively on the United States, Gottschalk and Smeeding (1997) broaden the horizon to include various European and other countries. They pay considerably more attention to institutions and signal important problems of conceptualization and measurement and stress that both market forces and institutional constraints cannot be missed in the analysis. Gottschalk and Joyce (1998) focus on the evolution of inequality in international comparison and conclude that market forces can be used to explain much of the cross-national differences that have been attributed in the literature to differences in LMIs. They hasten to add that this does not mean that institutional explanations do not matter but that the presumption should not be that they always provide binding constraints. Basing themselves on both direct and indirect evidence, Katz and Autor (1999, Section 5.5), in their comprehensive overview of the literature on wage structure and earnings inequality, view SBTC as perhaps the most important driver of the long-run growth in demand for more educated workers. They are less assertive about an acceleration in the trend. One important issue for further research they see is that the advantage of the better skilled may be transitory and for the long run depend on a continuing chain of technological changes or, alternatively, that 20th-century technological changes may happen to be systematically skill biased. Another question is whether the change is exogenous or may endogenously be affected by the supply of different skills.

Parallel to the market view, the institution-focused approach thrives. Fortin and Lemieux (1997) convincingly demonstrate its relevance basing themselves on both current American experience and effects of the Great Depression. There are some empirically obvious candidates for explaining wage inequality—the extent of union membership and the minimum wage given the clear declines in both the United States and the significant international differences, including the (de)centralized nature of wage bargaining. Freeman (1991) finds an important but not overwhelming role for unionization in relation to increasing inequality; by contrast, DiNardo and Lemieux (1997) explain two-thirds of the American-Canadian difference in male wage inequality growth from the faster decline of American unions; Blackburn et al. (1990) attribute only a small role to the declining minimum wage. DiNardo et al. (1996) find a substantial contribution to increasing wage inequality for the declining level of the minimum wage between 1879 and 1988. Though Levy and Murnane (1992) mention these institutions in their overview, they barely touch upon institutions—and then for the United States only—and apparently see no clear research agenda there. During the rest of the 1990s, however, the interest in institutions widens far beyond the above-mentioned. This receives great stimulus from international comparisons where differences in institutions can more easily get into the limelight. *Working under Different Rules* (Freeman, 1994), *Differences and Changes in Wage Structures* (Freeman and Katz, 1995), and Blau and Kahn (1996) bring together important empirical studies of a variety of countries. On a somewhat different tack from wage inequality, Freeman (1988) and Richard Layard and Stephen Nickell (1991) contribute on LMIs and macroeconomic performance in OECD countries. At the same time, Card and Krueger (1995a) dispel the consensus regarding the negative employment effects of the American minimum wage. As we will see later, these effects have remained a bone of contention up to this very day for their effects on employment, but much less for those on the wage distribution that are generally agreed to be compressing (Blau and Kahn, 1999). The *Handbook of Labour Economics*, Volume 3, concludes the 1990s by offering a rich palette of contributions on wage formation and inequality and institutions. Katz and Autor (1999) explicitly focus on earnings inequality and explanations from supply and demand on the one hand and from institutions on the other hand; Blau and Kahn (1999) treat LMIs in detail, and Brown (1999) specifically considers the minimum wage literature. Nickell and Layard (1999) also discuss the effects of LMIs but in relation to economic performance in general (the "natural rate of unemployment" and the later NAIRU), a different strand in the literature going back to Friedman (1968).

18.4.2 Defining and Analyzing LMIs

Given the importance of the subject for this chapter we will pay attention to the issue raised by the treatment of institutions before turning to the discussion of recent developments in the inequality literature in the next two sections. We distinguish between conceptualizing (method) and analyzing institutions and consider this at a general level in the present subsection, which boils down to the question of how economic analysis accounts for the existence of institutions. More specifically, their role in wage-inequality analysis is discussed in Section 18.4.5.

18.4.2.1 Method

First, the definition of institutions, and of LMIs in particular, seems in need of more precision, generically as well as in specific cases.[70] Too often in the literature they seem to be considered too obvious—and therewith perhaps too difficult—to warrant explicit definition. As Nickell and Layard (1999) say, "It is difficult to define precisely what we mean by labour market institutions, so we simply provide a list of those features of the labour market which we shall consider." In the words of Freeman (2007), "While economists do

[70] In the words of Freeman (2000, 11) "The absence of a general metric for measuring institutions at the national or firm level creates a problem for institutional economics. Measurement is, after all, the sine qua non of any scientific endeavour." See Autor (2013) for a similar worry regarding definitions and measurement for the tasks approach.

not have a single tight definition of an institution, per Justice Potter's famous statement about pornography, they know institutions when they see them, and they see them everywhere." Those who actually venture a definition come up with very broad ones, such as "A labour market institution is a system of laws, norms or conventions resulting from a collective choice and providing constraints or incentives that alter individual choices over labour and pay" (Boeri and Van Ours, 2008, 3), that may be begging important questions—in this case: Is the alteration of choice a practical matter or a theoretical issue? If the latter, this poses the risk that by definition any institution will imply a deviation from the theoretical ideal and also that institutions will be conceptually difficult to endogenize, meaning that an institution might endorse changes in behavior that have already taken place instead of causing such changes. It may also make the identification of institutions differ between different theories. Further to the restriction to laws, norms, or conventions: what about organizations? A trade union is an organization aimed at furthering the interests of its members.[71] Can a policy be an institution or not—as, e.g., active labor market policies are regarded an institution by Nickell and Layard (1999) and by Eichhorst et al. (2008)?

Institutions started their career in the inequality discussion as union density and the (de)centralized nature of wage bargaining. These are factors that can obviously influence individual wages, but they are neither law, nor regulation or rule, and seem closer to physical organization. The minimum wage, another institution also present from the start, is (often) in the law though. Not only in the inequality debate but more broadly, institutions have come to encompass a wide array of factors, and there seems to be no clear defining limit as to what can qualify as one or not. This makes institutions not only difficult to delineate but also tends to lend them a fundamentally *ad hoc* character. The selection of LMIs does not necessarily result from a systematic scrutiny but seems to reflect trial and error based on constrained and sometimes even biased knowledge. How otherwise to explain the immensely strong focus on unions while employer associations hardly figure, in spite of the fact that they are equally involved in collective bargaining and that "employer density"—the percentage of workers in a sector who are employed by the negotiating firms—and not union density may actually provide the basis for declaring a collective agreement generally binding.[72] The attention paid first to union density and only much later, when this appears to fail as an explanation, to the coverage of

[71] Checchi and Garcia Peñalosa (2008, 607) do distinguish "employee organizations outside the direct control of policy-makers" as one form of "collective intervention" encompassed in LMIs.

[72] Alan Manning (2011, 978–979) draws attention to employer collusion and points to research showing that "some institutions and laws in the labour market serve to aid collusion of employers to hold down wages." Though, in his view, it is clear that employers do not *en masse* collude, it would still seem logical to test centralized wage bargaining that results in protracted wage moderation for its potential as a form of nationwide monopsony.

collective bargaining, is another case in point.[73] Last but not least, it is difficult to understand, given the highly central role of educational differentials in the wage-inequality literature, that the educational system figures so little as an institution in this literature.[74]

In addition, in the literature, LMIs can be viewed as specific factors with an actual origin and presence in the labor market and only there (e.g., wage bargaining), but equally they may be factors that affect the labor market from outside that market (e.g., income taxation or the tax wedge). The latter definition, as a factor affecting the labor market irrespective of its origin, opens the door to a myriad of institutions and, unsurprisingly, many other factors have been added in the 30-year course of the debate. For example, Nickell and Layard (1999, 3047) include home ownership in their list of labor market "institutions" (their quotation marks), inspired by the finding of Oswald (1996) that it is one of the most important barriers to the geographical mobility of labor.[75] Freeman (2007) lists, without presumption of being exhaustive: mandated works councils, employment protection laws, minimum wages, extension of collective-bargaining coverage, lifetime employment, peak-level collective bargaining, wage flexibility, teams, job rotation, temporary employment contracts, social dialogue, apprenticeship programs, occupational health and safety rules, defined benefit, and defined contribution pension plans. To name a few other examples: Oliver (2008) draws attention to industrywide wage scales, Boeri (2011, 1183) adds regulations on working hours, Blau and Kahn (2002, 4) include the public-sector share of employment. Antidiscrimination measures easily come to mind as a further example,[76] and we will meet more below when we discuss recent contributions to the literature. Obviously, the discussion of earnings in relation to household income in the preceding section suggests new candidate institutions such as parental leave, child care provisions, or individual entitlements to the choice of working hours.

Second, the analysis of the effects of institutions still demand attention even if they were clearly defined, for several reasons: the actual significance of individual institutions and the type of effects they may have, their embedding in a larger set of institutions and also in the wider economy, and the potential pitfalls of international comparisons which have taken center stage in the literature.

[73] Availability of internationally comparative data was also a problem (e.g., Koeniger et al., 2007, 344).

[74] Leuven et al. (1997, 2004) criticize deriving skill levels from, inter alia, years of schooling in international comparisons. Nickell and Layard (1999, 3046) point to the same. Freeman and Schettkat (2001) and Mühlau and Horgan (2001) elaborate on the issue for the low skilled. Based on data from the International Adult Literacy Survey of the 1990s these contributions point out that the American low-skilled attain much lower levels of literary and numeracy than their counterparts in various European countries. At first glance the recent Survey of Adult Skills seems to underline the same 29 years later (OECD, 2013, 118–125).

[75] See also Blanchflower and Oswald (2013).

[76] Charles and Guryan (2007, 2008) explain one-quarter of the racial wage gap for blacks in the United States from employer prejudice.

In many studies institutions seem to be taken at face value and equated with what they look like *de jur*, and, naturally, it is their *de facto* implementary force or "bite" that counts (Eichhorst et al., 2008, 18). That bite may depend on its particular enforcement—laws and rules can be strictly enforced or they may be a dead letter to which no one pays attention. Enforcement may be automatic when it is the responsibility of a supervising inspectorate, or it may be costly and cumbersome if it is the responsibility of the individual who feels duped—minimum wages and other provisions diverge importantly in this respect (Benassi, 2011). Institutions may also be general provisions whose precise nature is filled in by actual policy. The minimum wage is again a case in point when the law establishes its mere existence while its actual (uprated) level is determined by policy making—the United States being the leading example.[77] For policies the bite is the heart of the matter.[78] Note that this further blurs the distinction between institution and policy. Institutions may also differ in the nature of their implementation: legally prohibiting or prescribing certain behavior, or economically encouraging or discouraging it, and consequently in the type of effects they may have. An example may be a prescriptive rule of employment protection versus a hiring subsidy for disadvantaged groups of active labor market policies. Note that the implementation of an institution may not be either black or white but can have different shades, a cross section within as well as between countries, and can also differ over time.[79] Evidently, the *de facto* significance of an institution may come close to actually measuring its effects, and this can pose a methodological problem.

Often the effects are scrutinized for *individual* institutions as there is no clear theory on their coherence or interactions as a (national) set (Eichhorst et al., 2008, 17, 24, 29). However, institutions may partly balance or reinforce each other, say a country with strong employment protection could be mitigating the possible upward pressure on wages by collective bargaining, or employment "at will" may be neutralized by individual contracts. It is easier to expand union membership if workers are protected from the threat of being fired. Diffusion of part-timers may reduce the quest of work leave permits. Generosity of unemployment benefit schemes may increase voluntary mobility and raise the demand for publicly provided training. Similar functions may be provided by different institutions. For example, Garnero et al. (2013) conclude to the functional

[77] See Boeri (2012) for a comparison of the effects of minimum-wage setting mechanisms, depending on the roles of unions and employers and the government in the process, across 66 countries.

[78] For the bite of the minimum wage see Kampelmann et al. (2013, 12–16) who discuss its relative money level in the wage distribution (Kaitz index) together with the share of minimum-wage employment in total employment. Eurostat's database on minimum wages ([earn_mw_cur] at http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home) is restricted to the former.

[79] For example, for involuntary dismissals, Dutch employers have the choice between following an administrative procedure, with no costs apart from the time it may take to settle, or going to court, which normally will be costly as it implies a severance payment. In recent decades the choice between the two routes has drastically shifted toward the latter option even though the administrative procedure has become more efficient (Salverda, 2008, 105).

equivalence with regard to earnings inequality of statutory minimum wages in some European countries and minimum-pay provisions of sector collective labor agreements combined with high bargaining coverage in other countries.[80] Gottschalk and Smeeding (1997, 647) warn for the risk of double counting the effects of institutions when considered in isolation (union density and the minimum wage in their example).

There may be deeper dangers in international comparisons. Institutions may catch the eye more readily than other international differences and be considered in relative isolation, enhancing the risk of their effects being overestimated. Blau and Kahn (1999) focus their overview on some 20 OECD countries stating that this selection permits utilizing "the similarity in educational levels, technology, living standards and cultures among these countries as de facto controls in examining the effects of institutions." Freeman (2007, 18) cautions against the methodological implications of the fact that the number of countries is small compared to that of institutions. Conversely, appearances may be deceptive and the potential dissimilarities of institutions that may look the same at first sight, need to be taken into account. We have already seen the possible divergence between educational attainment and skill levels in spite of the extensive efforts spent on a standardized measurement of educational systems (ISCED). Freeman's (2000) plea for a metric may be more demanding than thought but above all it may be necessary but not sufficient. The above observations about the enforcement and bite of institutions apply particularly in a comparative context, to prevent comparing apples with pears. More importantly, there may be deep-seated differences in the general economy as have been illustrated forcefully in recent years by, for example, the havoc wreaked by the larger propensity to consume by private households in the United States as compared to many other countries.[81]

In sum, the study of the role of institutions needs to account for the force of those institutions, their mutual interactions at the national level, supply and demand in the labor market, and also the broad structure of the economy. The latter potentially puts on the research agenda institutions that affect the economy more broadly such as those governing the flexibility of exchange rates[82] or international capital movements, which have undergone important liberalization in many European countries since the end of the 1970s and may have weakened employees and unions vis-à-vis employers. An important lesson of the minimum wage debate and the contribution made by Card and Krueger (1995a), who no longer started from the *a priori* of a negative effect on employment, is to prevent a stacking of the cards against institutions. As Freeman (2007, 2) observes, "many adherents to the

[80] The resulting levels may differ though.

[81] Glyn et al. (2003) find that the European–American services-employment gap resides largely in the distribution (retail) activities and personal-services sector and show "that the much lower European level of goods consumption per head of the population was the dominating influence in explaining the much lower levels of employment than in the US distribution" (p. 173).

[82] Blau and Kahn (1999, 1454) in their conclusions argue that exchange rates can adjust to compensate for institutional rigidities (and warn that introducing the euro may take away that opportunity, which has been borne out in the meantime).

claim (that labour institutions impair aggregate performance, authors) hold strong priors that labour markets operate nearly perfectly in the absence of institutions and let their priors dictate their modelling choices and interpretation of empirical results."

18.4.3 Why Do LMIs Exist?

The economic rationale and potentially beneficial effects of creating and/or preserving LMIs need to be accounted for from the start.[83] According to Freeman (2007) there are three ways in which institutions affect economic performance: by altering incentives, by facilitating efficient bargaining, and by increasing information, communication, and possibly trust. In his cursory review, the evidence shows that labor institutions reduce the dispersion of earnings and income inequality, which alters incentives, but finds controversial effects on other aggregate outcomes, such as employment and unemployment.[84] In his opinion, the modest effect would be attributable to the fact that "the political economy of institutional interventions rules out collective bargaining settlements and regulations that are truly expensive to an economy. No country would impose a minimum wage that disemployed a large fraction of the work force; and no union or employer would sign a collective bargaining agreement that forced the firm to close." In this perspective, a positive contribution of institutions would be observed whenever and wherever they solve transaction cost of individual bargaining, according to the prediction of the Coase theorem.

Regulations in the labor market as defined by Botero and coauthors (2004) emerge by government desires to protect the weaker side in a labor relationship.[85] They show that

[83] Some recent examples are Acemoglu (2003), who argues that LMIs stimulating wage compression in Europe may also incentivize investments in improving the productivity of the low skilled; Sutch (2010), who points to capital deepening and increased educational attainment as a consequence of the minimum wage (compare also Freeman, 1988); Nickell and Layard (1999), who consider that employee representation rights may induce management/worker cooperation and enhance productivity; or Atkinson (1999), who demonstrates that unemployment benefits, if accounting for their real-world rules, may actually be employment enhancing. A more fundamental, long-run perspective relates the origins of the welfare state, be it Beveridgean or Bismarckian, to the development of dependent employment.

[84] Similarly, Betcherman (2012) indicates four rationales for the existence of LMIs: "imperfect information, uneven market power (between employers and workers), discrimination, and inadequacies of the market to provide insurance for employment-related risks" (p. 2). According to him, the literature can be classified according to a positive view (that he calls *institutionalist*), when institutions solve coordination problems, and a negative view, that he calls *distortionary*, when institutions prevent economic efficiency.

[85] "Regulation of labour markets aiming to protect workers from employers takes four forms. First, governments forbid discrimination in the labour market and endow the workers with some 'basic rights' in the on-going employment relationships, such as maternity leaves or the minimum wage. Second, governments regulate employment relationships by, for example, restricting the range of feasible contracts and raising the costs of both laying off workers and increasing hours of work. Third, in response to the power of employers against workers, governments empower labour unions to represent workers collectively, and protect particular union strategies in negotiations with employers. Finally, governments themselves provide social insurance against unemployment, old age, disability, sickness and health, or death" (Botero et al., 2004, p. 1342).

the orientation of governments to the political left is often associated to more stringent labor market regulations (*political power theory*), but find the legal origin to be even more relevant to accounting for cross-country variation (especially when considering the transplantation of legal systems in the colonial era, much in line with sociological theories of path dependence—*legal origin theory*). According to the latter, common-law countries tend to rely more on markets and contracts, whereas civil-law (and socialist) countries on regulation (and state ownership): as a consequence, civil-law countries do regulate labor market more extensively than common law ones. The legal origin, possibly adopted for efficiency reasons in mother countries, becomes exogenous for former colonies, thus allowing a study of its causal impact on the origin of institutions.[86] Following this line of argument, several papers account for the endogenous emergence of some LMI as an (optimal) solution for at least of a subset of agents. A controversial contribution to this approach is given by Saint-Paul (2000), who aims to identify gainers and losers of a given institution. In his view, each institution creates a rent (i.e., a difference between the paid wage and the outside option), which is unevenly distributed in the workforce. Because the employed workers enjoy most of the benefits of these rents, they obviously represent the largest constituency advocating the preservation of institutions (*political insider mechanism*). This has to be traded-off against the rise of unemployment, which is associated to higher wages, and this represents the most serious threat to the continuation over time of an institutional setup. If we accept Saint-Paul's view that the most relevant conflict within the workforce is between the skilled and the unskilled, then "labour market rigidities mostly redistribute between skilled and unskilled labour" (Saint-Paul, 2000, p. 6). Ignoring within-group inequality, this means that institutions affect earnings inequality by affecting the skill premium and the (unskilled) unemployment rate. In this perspective, institutions emerge when the constituency represented by the employed unskilled dominates those of the skilled and of the unemployed (which is a different coalition than the one supporting fiscal redistribution, for example). The relationship between inequality and institutions becomes ambiguous: institutions create or enhance wage differences, but wage inequality may support the introduction of LMI as an alternative device for redistribution.[87] Similarly, different institutions may reinforce each other, revealing

[86] When they analyze the causal impact of legal indices on a measure of the skilled/unskilled differential, they find than only the "Social Security laws index" as an inequality enhancing causal impact (using the legal origin as instrument).

[87] "Inequality, i.e. the gap between the skilled and unskilled productivities, determines the intensity of internal conflict. As we have argued, it is because of that internal conflict that it pays the middle class coalition to opt for rigid LMIs. Therefore we expect that the support for rents will be greater, the greater the inequality. This is actually true over some range, if inequality is low enough. But past a certain threshold inequality reduces the support for rents, because at high inequality levels the cost of rigidity in terms of job loss is too big" (Saint-Paul, 2000, p. 8). See also Brügemann, 2012, who builds a model where stringent protection in the past actually reduces support for employment protection today.

the potential existence of a *politicoeconomic complementarity*, which contributes to explaining why empirically we observe clusters of institutions, often indicated as *social models* (Amable, 2003; Hall and Soskice, 2001). For reasons of viability, labor market reforms are more likely to emerge after a period of crisis, when the bias toward the status quo is weakened and the rise of unemployment allows the formation of alternative constituencies. A rather different view on rents in the labor contract, however, is offered by Manning (2011). According to him, rents are pervasive in the labor market, because of frictions in hiring and recruiting, separation costs due to investment in specific human capital, and collusive behavior on both sides (employers and employees). If imperfect competition is therefore taken as the relevant paradigm,[88] the regulation of this market (via wage bargaining or wage setting by public authorities, as in the case of minimum wage) acts as a second best device, which may achieve Pareto improvements (as in the case of the minimum wage under monopsony).

In a recent contribution, Aghion et al. (2011) frame the existence of labor market regulations as an incomplete and less-efficient substitute for the quality of labor relations. They rationalize their argument with a model of learning of the quality of labor relations: the unionization decision is seen as a costly experimentation device aimed at finding out more about cooperation at the workplace. Thus, the existence of legal provisions (such as a minimum wage) reduces the learning incentive. Because beliefs are gradually updated based on past experiences, the authors obtain the prediction of a coevolution of beliefs (as measured by the quality of labor relations perceived by top executives) and institutions (as measured by the stringency of minimum wage).[89] As a consequence, distrustful labor relations lead to low unionization and high demand for a direct state regulation of wages. In turn, state regulation crowds out the possibility for workers to experiment with negotiating and to learn about the potential cooperative nature of labor relations. This crowding-out effect can give rise to multiple equilibria: a "good" equilibrium characterized by cooperative labor relations and high-union density, leading to low-state regulation (the Nordic countries), and a "bad" equilibrium, characterized by distrustful labor relations, low-union density, and strong state regulation of the minimum wage (some of the Mediterranean countries, and especially France). Their empirical application covers 23 countries over the period 1980–2003, and shows that the quality of labor relations is negative correlated with either union density or state regulation of the minimum wage (while controlling for other institutional measures such as unemployment benefits and the tax wedge).

[88] "Many empirical observations (e.g., equilibrium wage dispersion, the gender pay gap, the effect of minimum wages on employment, employers paying for general training, costs of job loss for workers with no specific skills to list only a few) that are puzzles if one thinks the labour market is perfectly competitive are simply what one might expect if one thinks the labour market is characterized by pervasive imperfect competition" (Manning, 2011, 62).

[89] Although this prevents any causality analysis, it resembles the path dependence often advocated by sociologists in the analysis of institutions.

Also recently, Alesina et al. (2010) have proposed a model where the emergence of employment protection and minimum wage provision is accounted for by cultural traits, namely the strength of family ties. In their theoretical model, individuals are born with different preferences with respect to family ties: those characterized by weak ties are geographically or sectorally mobile and achieve an efficient allocation by being matched to jobs providing the highest productivity; however, those characterized by strong ties rationally select labor market rules (such as firing restrictions and minimum wages) that restrain the monopsonistic power of local employers while accepting a less-productive allocation. Another rationale of living close to family members is that it provides additional insurance against unforeseeable shocks (including unemployment). The authors prove the existence of two stable Nash equilibria: one where everybody chooses weak family ties, votes for labor market flexibility, and changes her or his initial location (high mobility); another where everyone chooses strong family ties, votes for stringent labor market regulation, and stays in the original (birth) location. In the latter case the labor market is monopsonistic because workers are immobile, and workers limit employers' power by means of labor regulations. Empirically, they show the existence of positive cross-country correlations between the strength of family ties and labor market rigidities. More convincingly, they also find that individuals who inherit stronger family ties (i.e., second-generation immigrants from countries that record high preferences for family values) are less mobile, have lower wages, are less often employed, and support more stringent labor market regulations.

In their historical review of the introduction of severance payment schemes in 183 countries, Holzmann et al. (2011) suggest three rationales for the introduction of such schemes: (1) as a primitive form of social benefits (anticipating the introduction of benefits for unemployment and retirement), thus providing an *answer to a demand for insurance*; (2) as an efficiency-enhancing human resource instrument (a sort of bonding between workers and firms, to minimize the loss of firm-specific knowledge) solving the *hold-up problem*; and (3) as a proper job-protection instrument, intended to enhance permanence in employment of main earners in the household.

If we restrict ourselves to the minimum wage, the historical account provided by Neumark and Wascher (2008) suggests that this institution has emerged as a counterbalance of power in the labor contract, preventing the exploitation of child labor (minimum-wage settlement power assigned to law courts in New Zealand in 1894 and in Australia 2 years later) or women (Fair Labor Standards Act, introduced at the federal level in the United States in 1938). Viewed in this perspective, the minimum wage would represent a device aimed at preventing a "race to the bottom" competition among firms, more than a measure aimed at sustaining the incomes of poor families.[90] Seen from

[90] In a similar vein, Agell and Lommerud (1993) proposed a model where setting higher wages promoted higher growth by eliminating low productivity enterprises.

the side of union leaders, minimum-wage legislation represents an improvement in the outside options of workers, inducing an increase in their bargaining power. The sum of these two effects may create an unusual coalition of large companies and worker unions supporting the introduction and/or the periodical updating of wage minima.[91]

18.4.4 Do LMIs Matter for the Economy?

In their overview of LMIs, Blau and Kahn (1999), with a careful discussion of the rationale, draw some implications for studying the causalities. They look back at "an explosion of research" on the economic impact of institutions and conclude that institutions do appear to matter. In their view, the evidence across the literature that institutions affect the distribution of wages is more robust than for employment levels. Freeman (2001) supports this, saying that institutions identifiably affect the distribution, but that other effects on the macroeconomy and on efficiency are hard to discover and modest at best. Later he states even more forcefully that "institutions have a major impact on one important outcome: the distribution of income . . . By contrast, despite considerable effort, researchers have not pinned down the effects, if any, of institutions on other aggregate economic outcomes, such as unemployment and employment" (Freeman, 2009, pp. 19–20; see also Freeman, 2005). Nickell and Layard (1999, p. 3078) seem more reticent about the role of institutions when they conclude that "[m]ost of the gross features of unemployment and wage distributions across the OECD in recent years seem explicable by supply and demand shifts and the role required of special institutional features such as unions and minimum wages is correspondingly minimal." These are not the last words about the role of institutions with regard to the dispersion of wages—let alone that of earnings incorporating the hours dimension which we deem of special interest here—as we will see when we turn to more recent contributions to the literature in the next two sections and to our empirical approach in Section 18.5.

So over the 1980s and 1990s a vast literature has grown, which seems to tend into two main directions: supply and demand on the one hand, institutions on the other. Each side acknowledges the relevance of the other, there is talk even of an SDI (supply-demand-institutions) model (Freeman and Katz, 1995; Katz and Autor, 1999; see also Lemieux, 2010) but little has grown out of that since, and in reality—understandably given the above-mentioned concerns—the prime focus of the market view and the institutional view seem to have grown more independent of each other. The flurry of institutions make them look overdetermined, and, by comparison, technological change—the driver of supply and demand—underdetermined. Over the 2000s many new arguments have been developed: polarization of the distribution, offshoring of productive activities, sharp growth in the upper tail of the distribution, top taxation, focus on tasks and skills, two-tier

[91] See their review of empirical evidence based on minimum wage voting across US states (Neumark and Wascher, 2008, chap. 8).

nature of reforms of institutions, growing importance of performance pay, rise of "new institutions," and, last but not least, new contributions have been made with regard to the minimum wage. These contributions seem firmly placed in either one or the other of the two main directions. In this respect the recent Volume 4 of the *Handbook of Labor Economics* repeats the preceding Volume 3. Acemoglu and Autor (2011) hardly even touch upon institutions in their conclusions, whereas Boeri (2011) focuses exclusively on aspects of institutions.

We think that Manning's (2011) approach of imperfect competition in the labor market, which aims to leave behind the thinking in terms of canonical models and departures from these, may indicate a third route that can provide a different and ultimately more unified perspective. From the starting point that rents are inevitable and pervasive—though it is unclear how large they are and who gets them—Manning (p. 996) suggests that their very existence creates a "breathing space" in the determination of wages and allows the observed multiplicity of institutions on efficiency grounds. He concludes (p. 1031) that "[o]ne's views of the likely effects of labour market regulation should be substantially altered once one recognizes the existence of imperfect competition."[92] An important corollary seems that institutions do not "cause the labour market to function differently from a spot market" (Blau and Kahn, 1999, p. 1400) but that this market should not be considered a spot market but instead needs institutions for its proper functioning from the very start. Thus, a better principle for analyzing supply and demand as well as institutions may be that institutions are equally pervasive: every act of supply and demand goes together with an institution of some kind, and their existence and effects shall be accounted for from the start.

18.4.5 Recent Theories Based on Demand and Supply of Labor Inputs

The review of theories of earnings inequality provided by Neal and Rosen (2000) a decade ago focused on the allocation of workers to jobs (the Roy model), on individual human capital accumulation (the Ben Porath model), on the search models (yielding variations in tenure—for a recent review see Rogerson et al., 2005 or Rogerson and Shimer, 2011), and on imperfect observability of either ability or effort (efficiency wage and contract theories). They adopted an individual perspective of wage determination, which did not allow great scope for the institutional framework to affect the resulting earnings distribution. In such a perspective, wage inequality can be considered as the outcome of changes in the relative demand and supply of labor inputs. Starting from the original paper by Katz and Murphy (1992) and the literature originated since then (reviewed in Katz and Autor, 1999), the so-called *canonical model* predicts that the wage differential between skilled and unskilled workers accommodates an expanding demand for skilled

[92] Note, however, his observation that the actual effects of (or, for that matter, the limits to) institutions are an empirical matter.

labor (SBTC, induced by introduction of computers in production) and a contraction of the demand for unskilled labor (due to increasing competition by developing countries). Demographic changes (variations in cohort size, immigration) and/or educational choices may partly attenuate (or even offset) these changes. The resulting dynamics of inequality can be predicted by tracking down these movements (Acemoglu, 2003).

In this framework, wage-setting institutions affect the flexibility of relative wages, creating a trade-off between wage differential and relative unemployment; when considering inter-industry wage differentials, it translates into lower employee quit rates and longer queues of job applicants.[93] Consider, for example, an increase in the relative demand for skilled labor (*upskilling*), at given supply of labor inputs. If the wage differential cannot adjust the relative excess demand for skilled labor (because minimum-wage legislation prevents a downfall of the unskilled wage and/or union bargaining prevents an excessive rise of the skill premium), then the unskilled workers will experience an increase in their relative unemployment rate. This effect will be more pronounced the higher the substitutability between labor types.

It did not take long into the new century before Card and DiNardo (2002) mounted a fierce critique of the thesis of skill-biased technological change. Their arguments are both theoretical and empirical. From a theoretical point of view, a constant SBTC rate does not yield a permanent skilled/unskilled wage differential, as long as the relative supply is sufficiently elastic (see Atkinson, 2007b). On the empirical side, they revisit the evolution of American wage inequality since 1967, almost back to the starting point of the literature but now extending to include more recent occurrences over the 1990s. This refers to the problem already mentioned that technological change lacks a positive identification in economic models but is commonly subsumed in the unexplained leftovers. To avoid the tautology that this implies, they look for independent empirical measures of technological change that can be incorporated in the model: the introduction of PCs and the Internet, the size of the IT sector in the economy, and the use of computers by individuals at work—particularly disaggregated by personal characteristics.[94] From this material, the general trend in technological change seems unabated over the 1990s, if not increasing because of the Internet. The disaggregated use of computers points, among other things, to a larger role among women than men, particularly among the less-educated women whereas the best-educated men have closed the gap to their female counterparts. From this, Card and DiNardo conclude that computer technology should have widened gender differentials for the most highly educated and narrowed them for the least educated.

[93] Katz and Autor (1999) also consider product market regulation, in that it creates differences in sectoral rents, which are partly appropriated by wage bargaining, thus contributing to the overall wage inequality.

[94] Note that DiNardo and Pischke (1997) show robust wage differentials for the use of pencils (in Germany) and draw attention to the possible selection effect that office tools tend to be used more by higher-paid workers.

On the inequality side, they argue from a fresh inspection of the data (using different samples, sources, and inequality measures), "viewed from 2002" as they say, that there has been a pattern of a strong episodic rise in inequality in the 1980s, preceded by near stability before and after, during the 1970s and 1990s respectively.[95] From a comparison of the two, demand and supply, they conclude to "a fundamental problem . . . that rises in overall wage inequality have not persisted in the 1990s" and also to various puzzles, including the fact that the gender differential has diminished irrespective of education. In summary, they find the evidence for SBTC to be surprisingly weak. They do think there has been substantial technological change but deplore that this has diverted attention away from inequality trends that cannot be easily explained by this. The critique of SBTC is the main point of their contribution, not the design of an alternative explanation of inequality. However, Atkinson (2007b, 2008) points out that their critique of SBTC ignores the dynamics of the process and implicitly assumes a curve of skilled labor supply whose speed of adjustment is inversely related to the distance from an infinitely elastic one. International differences in the wage differential may reflect differences in the speed of that adjustment. Card and DiNardo end their contribution by teasing the reader with a quick exercise about the minimum wage that shows a strong correlation between the evolution of its real level and aggregate hourly wage inequality (P90:P10) over the entire period 1970–1999.

Autor et al. (2006, 2008) have shown that the period of rising earnings inequality in the US labor market during the 1970s and the 1980s has been replaced by *job polarization* (simultaneous growth of the share of employment in high-skill/high-wage occupations and low-skill/low-wage occupations) in the following two decades. Despite the fact that the emergence of polarization crucially hinges on the procedure according to which occupations are ranked (educational attainment, wage rank, task content), also many European countries feature similar patterns: the decline in blue-collar jobs (mostly held by uneducated men) and the expansion of service jobs (mostly held by women and youngsters). One suggested interpretation (Autor et al., 2003) points to the increase in productivity of information and communications technology (ICT), which would have replaced middle-skilled administrative, clerical, and productive tasks with computer-operated machines.

Autor et al. (2008) have taken up the challenge of what they call a "revisionist" literature of both the description and the explanation of US wage inequality since the 1970s. They object to the episodic interpretation of the rise in wage inequality; that is, they contrast this with ongoing inequality growth in the top half of the distribution combined with initially (1980s) increasing and subsequently (1990s) declining inequality

[95] Lemieux (2006a,b) finds a concentration of the increase in the 1980s together with a concentration of within-group inequality change among male and female college graduates and females with some college, implying an increasing concentration of wage inequality at the very top of the wage distribution. In addition, Lemieux (2006c) finds a role for changes in the composition of the labor force after the 1980s.

in the bottom half.[96] They view that initial lower-half increase as episodic indeed and incorporate the minimum wage as a potential explanatory factor in their approach; however, they find only a modest role when modeled together with relative supply and demand. For the 1990s they agree that the slowing down of inequality growth poses a problem for the SBTC thesis, but only for the "naïve" SBTC story as they call it, which is based on a dichotomy of high skills and low skills. They aim to improve on this by arguing a more detailed approach, based on the dispersion of occupations by their skill levels, measured as the mean years of schooling of an occupation's occupants (weighted by their hours worked), and distinguishing between different types of tasks that can be performed in an occupation, showing that this works out differently between the 1980s and the 1990s.

The occupations and tasks approach can be viewed as a step along the route for further research pointed out by Levy and Murnane (1992), opening up an important black box albeit at the level of industry and not of the firm.[97] In principle, though not always in practice, it also advances on the traditional SBTC approach by distinguishing between properties of the occupation and of the worker. Routine tasks were first stressed by Autor et al. (2003), polarization by Goos and Manning (2003, 2007). The approach aims to provide an answer to the problem posed to the SBTC thesis by the strong slowdown in wage inequality growth after the 1980s. It implies a significant shift in the SBTC thesis and the underlying empirics. Modern technology is complementary no longer to higher levels of skills and education but to nonroutine types of work. Although before workplace computerization was indiscriminately interpreted as skill biased and furthering the demand for higher skills, it is now taken to substitute for routine tasks that are defined as cognitive and manual activities that can be accomplished by following explicit rules. Therewith it reduces the demand for workers predominantly performing such activities, implying a more polarized effect on educational levels. Autor et al. (2003) focus on American employees, and the period 1960–1998 and combine CPS data with Dictionary of Occupational Titles (DOT) classifications. They analyze the shift in tasks that has resulted from both compositional changes across occupations and changes in task composition within occupations, and find strongly diverging trends: negative for routine cognitive tasks from the 1970s and routine manual tasks from the 1980s and strongly positive for nonroutine cognitive tasks whereas nonroutine manual tasks decline steadily and strongly over the entire period.[98] Note that they focus on employment effects and do not link the results to wage inequality,[99] though the implication is clear and to some extent spelled out in Autor et al. (2006): low-wage and high-wage employment both

[96] They agree to compositional effects (Lemieux, 2006a), but for the lower half of the distribution only.

[97] Dunne et al. (2004) find most of the action between and not within establishments in US manufacturing.

[98] The authors consider the latter tasks as orthogonal to computerization and therefore not impinging on their results.

[99] Interestingly, Goos and Manning (2007) fill that gap, showing—for 1983 only, the first possible year—that routine jobs are concentrated in the middle of the US wage distribution.

expand while jobs with intermediate pay contract. Notably, employment trends do not seem to differ between the 1980s and 1990s, though, naturally, the gaps between increasing and decreasing types of tasks become much wider.

Goos and Manning scrutinize the UK data for similar developments between the mid-1970s and the late 1990s, using a variety of data sets, samples, and methodological approaches. They find a clear polarization across the distribution of occupations and, linking to wages, also across the wage distribution. They check various tenets of the SBTC thesis. They discuss first whether labor supply may have contributed to the polarization, because of the rapid growth in female workers, and better-educated workers, but they find these changes unable to explain the polarization pattern. As to educational attainment, they find an increase in almost all occupations. This may be due to either rising requirements of the jobs or overeducation of the occupants. The data are insufficient to decide between the two hypotheses though the authors seem inclined to opt for the second one. On the demand side they touch upon other factors than technology that may have contributed: trade and especially the structure of product demand—though these are not necessarily fully independent from technology—but find no explanation for polarization either. From a counterfactual exercise of the wage distribution over the 1975–1999 period restricted to changes in the occupational distribution only they conclude that polarization can explain large fractions of the rise in wage inequality (51% lower half, 79% upper half). They underline the important implication that the contribution of within-job inequality is minor. This contrasts sharply with established explanations in terms of education and age where most of the action is within groups, and they point out that the between/within conclusion is sensitive to the choice of controls included in the earnings function. They leave open the explanation of inequality change in the lower half of the distribution which may be due to imperfect competition, including institutional changes such as declines in unionization or the minimum wage. Goos et al. (2009) show a polarization of employment by occupations for 16 European countries between 1993 and 2006; Goos et al. (2010, 2011, 2014) extend the analysis to include relative wages and also capture effects of product demand, induced by a lowering of relative prices in industries with routine tasks, and institutions. They find that relative occupational wage movements in Europe are not strongly correlated with technology and offshoring, which may be due to wage-setting institutions, and therefore consider relative wages as being exogenous. They conclude that the thesis of routine jobs is the most important explanatory factor for increasing polarization, and product demand shifts across industries mitigate it.

Dustmann et al. (2009) find increasing wage inequality for Germany in the upper half of the distribution over the 1980s and 1990s.[100] This is attributed partly to composition

[100] Spitz-Oener (2006) looks at the employment side of occupational polarization in Germany over the 1980s and 1990s.

changes and largely to technological change, as occupations at the top grow faster. For the lower half they find increasing inequality only in the 1990s, not before. For this they suggest possible episodic explanations such as a decline in unionization and an inflow into the country of low-skilled labor after the demise of the communist regime; the latter lends a role to the relative supply of skills. Following Autor et al. (2008) they conclude that the naïve or canonical SBTC hypothesis cannot explain these trends, but they find support for the "nuanced" tasks-focused hypothesis as they note that occupations in the middle of the distribution decline compared to those at the bottom. In summary, they believe that the German results add unifying evidence to the pattern of polarizing effects of technological change already found for the United States and the United Kingdom.

We conclude our discussion of this stream of the literature with its current culminating point, the overview and further development of the task-based approach to SBTC by Acemoglu and Autor (2011) for the latest *Handbook of Labor Economics* (Ashenfelter and Card, 2011).[101] Note, however, that Mishel et al. (2013) provide various arguments why the evidence for the job polarization of these is weak. Although the canonical model builds on the unity of skills, tasks, and job (better-educated/talented workers obtain skilled jobs where they perform more complex tasks), the task-based approach considers a job as a collection of tasks, which can be executed by workers of different abilities, though at different level of productivity, and even by machinery. The empirical classification of tasks is still in its infancy; they are classified according to three attributes: routine, abstract, manual. "Offshorability," meaning that the performance of certain tasks is internationally footloose, is added as another important job dimension, which can overlap with each of the three types of tasks.[102] This theoretical approach improves upon the canonical model by accounting for job polarization, real wage decline for some groups of workers (but not in a monotonic relationship with skill ranks), and offshoring as an alternative explanation of reductions in jobs to technical change.

[101] Autor (2013) adds a further overview of the literature stressing the need to develop a precise terminology and consistent measurement. He honorably concludes that "[t]he economics profession is very far from a full understanding of the interactions among rising worker skills, advancing technology, improvements in offshoring and trade opportunities, and shifting consumer demands in determining the division of labour, the growth of aggregate productivity, and the level and inequality of earnings within and between skill groups. The 'task approach' to labour markets does not come close to offering a solution to this vast intellectual puzzle" (p. 27). Surprisingly, he also sounds an optimistic note about the future of middle-skill jobs. Autor and Dorn (2013) make a further addition venturing consumer preferences as a second force next to technological change that can help explain polarization through the growth of low-skill services in the United States. However, one cannot be sure about the general validity of this approach as consumer preferences may differ significantly across countries.

[102] The concept of offshorability of jobs and its analysis in relation to wage inequality was developed during the 2000s by Levy and Murnane (2005), harking back to Blinder (2007), Lemieux (2008), and Blinder and Krueger (2009). Evidently, offshorability itself is conditional on both technological change and institutional preconditions.

Acemoglu and Autor's model considers a continuum of tasks (unit of work activity that produces output, similarly to occupations) and different levels of skills (capability to perform various tasks); given existing supply of skills in the labor market, profit maximizing firms allocate skills to tasks, given existing prices. Capital and/or offshoring may replace workers in performing tasks. The key assumption is the existence of comparative advantage of skills in executing tasks: more skilled workers are more productive in executing more complex tasks when compared to less-skilled workers. This structure creates a sort of *hierarchical sorting* associated to comparative advantage.

Wage flexibility ensures full employment of all workers. Given perfect substitutability among workers in task assignment, wages dynamics depend on the relative supply of skills (as in the canonical model) and on task assignment rules, which then allow for a potential competition in task execution posed by technological progress and/or offshorability. With their model they make a sharp prediction: "[I]f the relative market price of the tasks in which a skill group holds comparative advantage declines (holding the schedule of comparative advantage constant), the relative wage of that skill group should also decline—<u>even if</u> the group reallocates its labour to a different set of tasks (i.e., due to the change in its comparative advantage)" (p. 1152). The impact on the overall wage inequality is hard to predict, because the relative wages (high to medium skill and medium to low skill, when only three skill levels are considered) can move in opposite directions.

Acemoglu and Autor do not incorporate LMIs in their framework, which as they observe "depends crucially on competitive labour markets" (p. 1159) and can be thwarted by labor market imperfections of search and information and institutions such as collective bargaining by unions. The impact of certain LMIs may be enhanced by the way these affect the assignment of tasks to labor or capital as, for example, they may restrict the substitution of machines for labor for certain tasks, or conversely they may change the return to unionization, thus feeding back onto union density. The authors see this as an area for further research.

18.4.6 Recent Theories Based on LMIs

The other main current in the literature does take the existence and effects of LMIs into account. Also here, interesting contributions have been made throughout the 2000s. At the start of the new century, Blanchard and Wolfers (2000) launched their hypothesis that the internationally differential effects of institutions can be found particularly in countries' responses to shocks. This view can offer a solution to the problem that, on the one hand, shocks alone cannot explain country differences, and, on the other hand, institutions on their own cannot explain long-run country performances. Their focus is the macroeconomy and unemployment, not wage inequality. Blau and Kahn (2002) connect to the latter in much of their book, and later extend this further by accounting for demographic shocks (with Bertola, 2007). However, their strong focus on the international comparison of institutions may be the reason that they seem to overlook the shifting trend in the evolution of American wage inequality after the 1980s. The issue of this shift has been taken up by Lemieux (2008), for the United States. He objects to the consensus view on inequality growth that had taken root in the early 1990s which views this growth as secular and all-pervading. As we have seen his contribution (2010) extensively revisits the American data on the evolution of wage inequality, and pays particular attention to the very top of the wage distribution, improving on the traditional adjustment for top-coding. From this, he concludes that in the 1970s inequality change was not all-pervading, whereas it was in the 1980s (though it also already showed more convexity at the top than at the bottom), and that since the 1990s inequality growth has been concentrated at the top of the distribution. Growth in residual ("within") wage inequality is general in the 1980s, although later it is largely confined to the college-educated category. In particular, relative wages continue to grow for postgraduates and their annual returns to education compared to high-school returns double between the mid-1970s and the mid-2000s.

Lemieux (2008) also questions the consensus explanation of SBTC on the basis of this, but also because it leaves no room for a role of institutions in spite of the research that has shown the effects of unionization and wage-setting. He advocates an explanation that can account for both the above findings and the international differences and explores the possible contributions of institutions as well as of supply and demand.[103] He finds that deunionization can explain one-third of the expanding inequality in each of the two halves of the distribution, and is also consistent with the divergence of English-speaking countries, where top incomes grew much more, from other countries. In addition to this the decline in the minimum wage has augmented lower-half inequality in the 1980s.[104] On the side of supply and demand he thinks that more empirical research is needed before the tasks-based development of the SBTC thesis can be accepted as an explanation. That research should account for the fact that, contrary to what one would expect, the relative wages of occupations at the core of the IT revolution are suffering, and it should also answer the question why the process should not have occurred already during the 1980s. In addition, it should account for the growth of within-inequality at the top. For the latter he suggests modeling heterogeneous returns to education, which have

[103] His main objection to SBTC is that technology is widely available across countries, whereas inequality growth is recorded only in the Anglo-Saxon world. However, similar impacts are now also recorded in developing countries (Behar, 2013).

[104] Lemieux et al. (2009) add, as an additional institution, performance pay at the top—bonuses, stock options, etc.—and show that this can account for a large share of inequality growth above the 80th percentile of the wage distribution.

as a key implication that both the level and the within-dispersion of pay of the better educated can rise relative to the less educated at the same time.[105]

18.4.6.1 Top Incomes

Interestingly, Lemieux's conclusion about upper-tail growth is consistent with the findings in the top-incomes literature (Alvaredo et al., 2013; Atkinson and Piketty, 2007, 2010; Atkinson et al., 2011, especially the summary Chapter 12; Piketty and Saez, 2003, 2006). Often a strong rise in labor incomes at the top is found, particularly in the United States, but not only there.[106] This literature is suggestive of the role of yet another institution: income taxation, not as the traditional tax wedge but as marginal taxation at the top.[107] "Higher top marginal tax rates can reduce top reported earnings through three main channels. First, top earners may work less and hence earn less—the classical supply side channel. Second, top earners may substitute taxable cash compensation with other forms of compensation such as non-taxable fringe benefits, deferred stock-option or pension compensation—the tax-shifting channel. Third, because the marginal productivity of top earners, such as top executives, is not perfectly observed, top earners might be able to increase their pay by exerting effort to influence corporate boards. High top tax rates might discourage such efforts aimed at extracting higher compensation" (Atkinson et al., 2011). Thus, the rise in top incomes and pay may have been encouraged by the lowering of top marginal tax rates. However causation may also run in the opposite way, because the rise of capital incomes in recent decades may have produced pressure for tax reductions. In a recent series of papers, (e.g. Piketty and Saez, 2013) have proposed formal models where the relationship between taxation and earnings has been carefully scrutinized. Most of the argument is a supply-side story, in the presence of imperfections: a reduction in the degree of progressivity would stimulate more effort and bargaining of CEOs and high-rank cadres with stakeholders, thus raising earnings inequality. Piketty et al. (2011) show a strong negative correlation between the Top 1% share and the top tax rate for a set of 18 OECD countries since 1960; the correlation also holds for CEO pay after controlling for firm characteristics and performance. The element of luck in CEO pay seems to be more important when tax rates are lower. It may point to more aggressive pay bargaining in a situation of lower tax rates. The high top tax rates of the 1960s were then part of the institutional setup putting a brake on top compensation through bargaining or rent extraction effects. In their view, the SBTC explanation seems to be at odds with international differences in top pay shares as well as their correlation to tax rates.

[105] Slonimczyk (2013) links overeducation to the differential growth of inequality in the two halves of the distribution.

[106] For example for the Netherlands (in spite of stability of the top income share as a whole): see Salverda and Atkinson (2007) and Salverda (2013).

[107] DiPrete (2007) highlights the increase in external recruitment of CEOs and the concomitant growth of related institutions (governance and CEO pay benchmarking).

18.4.6.2 Minimum Wage[108]

New contributions to the literature of inequality and institutions are also found for various other individual LMIs. First and foremost, we consider the literature on the effects of the minimum wage—an old debate by now (as old as the Department of Labor (viz. 1913) according to some)[109] that nevertheless continues to attract passionate contributions. The combination of wage and employment effects taken together determines the effects on annual earnings and, ultimately, incomes. Especially the impacts of a minimum wage on employment remain a bone of contention—"the canonical issue in wider debates about the pros and cons of regulating labour markets" in the words of Manning (2011, p. 1026). A complication is that the employment effects likely relate to the level of the minimum wage and also differ between worker categories (e.g., Abowd et al., 1999; and also Philippon, 2001).

Neumark and Wascher (2008) hold a very critical attitude with respect to minimum wages. Exploiting cross-state and temporal variations in the United States, they conclude that minimum wages are ineffective in raising low wages and reduce employment opportunities for their earners.[110] However, Dolton and Bondibene (2011) analyze employment effects for 33 OECD countries over 1976–2008 and find that existing evidence of negative effects is not robust. Dube et al. (2010) generalize Card and Krueger's comparison of minimum-wage policy differences across US state borders and find no employment effects over 1990–2006, whereas Neumark et al. (2013) dispute their method and results. Allegretto et al. (2011) find no employment effects (including the hours dimension) distinguishable from zero over 1990–2009. Slonimczyk and Skott (2012) use US state variation to confirm their model predictions of a negative effect

[108] There is also an emerging literature on developing-country case studies, which confirm the inequality-reducing impact of minimum wage, both in the formal and informal sectors of the economy (e.g., Gindling and Terrell, 2009; Lemos, 2009).

[109] Note that the UK minimum wage, introduced very recently in comparison with the United States (1999 vs. 1938) has been a great source of new evidence thanks to the careful role of the Low Pay Commission—see Butcher (2011).

[110] "Based on the extensive research we have done, and our reading of the research done by others, we arrive at the following four main conclusions regarding the outcomes that are central to policy debate about minimum wages. First, minimum wages reduce employment opportunities for less-skilled workers, especially those who are most directly affected by the minimum wage. Second, although minimum wages compress the wage distribution, because of employment and hours declines among those whose wages are most affected by minimum wage increases, a higher minimum wage tends to reduce rather than to increase the earnings of the lowest-skilled individuals. Third, minimum wages do not, on net, reduce poverty or otherwise help low-income families, but primarily redistribute income among low-income families and may increase poverty. Fourth, minimum wages appear to have adverse longer-run effects on wages and earnings, in part because they hinder the acquisition of human capital. The latter two sets of conclusions, relating to the effects of minimum wages on the income distribution and on skills, come largely from U.S. evidence; correspondingly, our conclusions apply most strongly to the evaluation of minimum wage policies in the United States" (Neumark and Wascher, 2008, p. 6).

of minimum wages on the skill premium, owing to increasing overeducation of college-educated workers following an increase in mismatch. The overall effect is that a minimum wage would lead to a rise in both total and low-skill employment, accompanied by a fall in earnings inequality. Giuliano (2013) studies personnel data of a large US retail firm and finds no aggregate employment effect but composition effects that run contrary to standard theory. Interestingly, Dube et al. (2012) focus attention on effects on employment flows. In the view of Richard Sutch (2010), the disemployment effects of pricing low-skill jobs out of the market may create incentives to invest more in human capital.

Most of the recent discussion revolves around whether there are spillover effects on wages higher than the minimum. A higher statutory minimum wage in itself compresses the wage distribution as it prohibits paying lower wages. However, the higher minimum rise may send ripples up the wage distribution—in the most extreme case all wages could be increased to the same extent and the dispersion of wages would remain unchanged. The minimum wage debate of the 2000s has generated new contributions particularly on this spillover or knock-on issue. Wages higher up may be raised for several reasons (Stewart, 2012, 618): the higher price for low-skilled labor incites substitution demand for higher-skilled workers, realignment of the marginal product of minimum wage workers affects the marginal product of other workers, firms maintain within-firm pay differentials for motivation, and reservation wages increase more broadly in certain sectors.

During the 1990s, spillover effects were detected in various contributions. Card and Krueger (1995a, 295) conclude to no effect at or above the 25th percentile of the wage distribution, which is well above the relative position of the minimum wage. Lee (1999) endorses an approach that compares to an estimated "latent" wage distribution (in the absence of the minimum wage). He finds effects beyond the P50:P10 ratio on other percentile differentials across the entire distribution. At the end of the 1990s, the consensus view agreed to spillover effects though not extending high up the distribution (Brown, 1999, p. 2149).[111] Over the 2000s, views on this have changed. Neumark and Wascher (2008, Section 4.3.2) discuss the previous literature and observe that the percentile approach as used by Lee may conflate spillover effects with disemployment effects of the minimum wage: as some of the least-paid lose their jobs, wage levels may increase at all percentiles of the distribution. Neumark et al. (2004) do not link to the wage distribution but look instead at actual impacts on workers with wages up to eight times above the minimum wage, using US states with no rise in their minimum wages as controls. They find a wage elasticity with respect to the minimum wage of 0.25 at 1.5 times the minimum wage and much smaller effects above that level. Autor et al. (2010) are puzzled by Lee's effects on the upper half of the distribution and attribute these to an omission of variables and the insertion of the median wage on both sides of the equation (division bias). They stick to Lee's basic approach but propose econometric corrections

[111] Lee (1999) is not covered in Brown's (1999) overview.

and demonstrate the effects using a longer panel of US states with more variation in state minimum wages. They find substantial widening effects on the lower tail (P50:P10) of the decline in the real minimum wage over 1979–1988, but these effects remain well below those found earlier in the literature; they find only small effects for 1988–2009. Then they are puzzled by the large and increasing effects even at the 10th percentile in spite of the fact that currently the minimum wage is received by less than 10% of workers. They confront those effects with the possibility of mismeasurement and misreporting of lower wages in the data and conclude from a detailed analysis that it cannot be ruled out that all of the spillover found is actually the result of such data problems.

Stewart (2012) adopts the direct estimations of Neumark et al. (2004) over a range of fractions of the minimum wage extending up to six times the minimum wage using differences-in-differences for comparisons between these factions. In addition, he exploits comparisons between minimum wage upratings that have differed in size (including no change period before the introduction of the minimum wage in 1999) while accounting for differences in general wage growth. Using British data, he concludes to no spillover effects. As the level of the minimum wage is steadily below the 10th percentile, he draws the logical inference that the changes in the minimum wage have not affected lower-half wage inequality as measured by P50:P10. That seems fair enough, but it also puts on the table the strength of this inequality measure as evidently the minimum wage may significantly affect the within-distribution of the bottom decile. The top-to-bottom ration S10: S1 may be better suited to capture such effects. Butcher et al. (2012) revisit the effects on wage inequality and spillovers for the United Kingdom and do find spillover effects up to the first quartile of the distribution. In their view, decades of discussing the employment effects of the minimum wage—with very little to none as the consensus outcome—have been focusing on second-order effect, and instead they advocate developing a theoretical framework for thinking about its first-order effects on wage inequality, which, naturally, should be able to allow the possible absence of employment effects. They develop a non-competitive model with wage-posting instead of bargaining[112] with imperfectly elastic labor supply to the individual firm. The authors elaborate on their model to consider the spillover effects to wage levels above the minimum wage. They derive those from a comparison between the actual wage distribution at and above the minimum wage and a counterfactual latent wage distribution derived with the help of the distribution preceding the introduction of the minimum wage in 1999. They find higher levels for the former compared to the latter up to 40% above the minimum wage, which corresponds with the 25th percentile of the aggregate wage distribution.

[112] Wage bargaining cannot explain the frequent uniform payment of the same low wage to workers with rather different characteristics. Hall and Krueger (2010, p. 25) conclude that their findings from a special survey of wage posting and bargaining practices in the US labor market “is consistent with the view that a wage constrained by the minimum wage is inherently posted.”

Finally, Garnero et al. (2013) show that statutory minimum wages (or equivalent systems) represented by sectoral minimum rates combined with high coverage of collective bargaining—see also Boeri (2012)—are very effective in reducing earnings inequality. They combine harmonized microdata from household surveys (EUSILC), data on national statutory minimum wages and coverage rates, and hand-collected information on minimum rates from more than 1100 sectoral-level agreements across 18 European countries over several years (2007–2009—see also Kampelmann et al., 2013). Alternative specifications confirm that institutional variants of setting a wage floor reduce both between and within-sectors wage inequalities.

18.4.6.3 Union Presence

Card et al. (2004) study the relationship between wage inequality and unionization in the United States, Canada, and the United Kingdom over the period 1980–2005, showing that within narrowly defined skill groups, wage inequality is always lower for union workers than for nonunion workers. For male workers, union coverage tends to be concentrated at the middle of the skill distribution, and union wages tend to be "flattened" relative to nonunion wages. As a result, unions have an equalizing effect on the dispersion of male wages across skill groups. For female workers, union coverage is concentrated near the top of the skill distribution, and there is no tendency for unions to flatten skill differentials across groups. The effect of deunionization on US wage inequality is stronger at the top end of the distribution than at the bottom, as shown by Lemieux (2008) when updating the DiNardo et al. (1996) decomposition. In addition, the increase of performance pay schemes may have enhanced the within-group wage inequality at the top end of US distribution.[113]

The decline in workers' bargaining power in the Anglo-Saxon world is recorded by several authors (see, for example, Levy and Temin, 2007), but we have not found any convincing decomposition of the relative contribution of each specific institutions. However, when taking the dynamics of the wage share in the domestic product as an overall indicator of workers' bargaining power, one would recognize a clear declining trend in most countries over the past decade, though some reversal can be recognized during the crisis period (ILO, 2008, 2010).[114]

[113] However, existing comparative evidence on differences in executive compensation between the United States and Europe suggests that this labor market is fully globalized, and pattern of remuneration are quite similar (except in the banking sector). See Conyon et al. (2011).

[114] "The slow growth in wages was accompanied by a decline in the share of GDP distributed to wages compared with profits. We estimate that every additional 1 per cent of annual growth of GDP has been associated on average with a 0.05 per cent decrease in the wage share. We also found that the wage share has declined faster in countries with a higher openness to international trade, possibly because openness places a lid on wage demands based on a fear of losing jobs to imports. Inequality among workers has also increased. Overall, more than two-thirds of the countries included in our sample experienced increases in wage inequality. This was both because top wages took off in some countries and because bottom wages fell relative to median wages in many other countries" (ILO, 2008, p. 59). See also Karabarbounis and Neiman (2013), who attribute the decline in wage share to the decline in the relative price of capital inputs.

A parallel decline in workers' bargaining power can underlie the decentralization of wage bargaining. Following recent changes in industrial relations in Denmark, Dahl et al. (2011) show the existence of a wage premium associated with firm-level bargaining relative to sector-level bargaining, and a higher return to skills under more decentralized wage-setting systems.[115]

18.4.6.4 Unemployment Benefit

Even if unemployment benefits and employment protection are negatively correlated in the data (Bertola and Boeri, 2003), in principle they do respond to the same problem of reducing the intertemporal variability of workers' earnings (Blanchard and Tirole, 2008).[116] This may explain why research has paid less attention to the contribution of unemployment schemes to inequality reduction. Corsini (2008) studies the dynamics of the college premium in 10 European countries over the last decade of previous century. He finds a positive impact of the generosity of unemployment benefit (but a negative correlation with duration), which is interpreted as the outcome of wage bargaining that takes into account the outside option.[117] If we shift to individual data analysis, the results of Paul Bingley et al. (2013) on Danish data show that access to unemployment insurance is associated with lower wage-growth heterogeneity over the life cycle and greater wage instability, changing the nature of wage inequality from permanent to transitory. Given data limitations, the authors are unable to control for moral hazard behavior of unemployed, who may be induced to lengthening their permanence in unemployment, thus increasing cross-sectional inequality.[118]

18.4.6.5 Employment Protection Legislation

Recent cross-country evidence has been summarized in the following way by World Bank (2012, 262): "Based on this wave of new research, the overall impact of EPL and minimum wages is smaller than the intensity of the debate would suggest." However, Martin and Scarpetta (2011) express a different view, arguing that EPL reduces workers' reallocation and prevents efficiency gains for highly productive workers, while avoiding

[115] Kenworthy (2001) discusses existing measures of wage-setting institutions.

[116] Chetty (2008) derives the optimal replacement rate for unemployment benefit schemes that depends on the reduced-form liquidity and moral hazard elasticities.

[117] Vroman (2007) discusses the correct measure of (average) unemployment compensation from aggregate public expenditure on subsidies, to be contrasted with standard OECD replacement rate and duration series, which are commonly used, despite their being completely hypothetical (because they are derived from microsimulation models) and do not correspond to actual payments to entitled unemployed workers.

[118] Using cyclical and across-US states variation, Farber and Valletta (2013) show that extending the duration of unemployment benefits (from a Federal requirement of a minimum of 26–99 weeks at the cyclical peak of late 2009) lengthens unemployment spells, via a reduction in exits from the labor force (and not in job finding due to reduced search effort).

job losses and/or real wage reductions for unskilled workers.[119] In their review they list a series of papers based on changes in dismissal regulation, which find mixed evidence of EPL impact on labor productivity (see, among others, Bassanini et al., 2009; Boeri and Jimeno, 2005; Kugler and Pica, 2008; Schivardi and Torrini, 2008). Productivity dynamics may translate one to one into wage dynamics in a competitive environment; in noncompetitive models, firing restrictions raise the bargaining power, creating artificial divisions among workers when groups of firms are exempted (see Leonardi and Pica, 2013). Similarly, EPL exemptions for firms may create artificial wage differences among workers, due to their differential cost, thus enhancing wage inequalities; for example, Karin Van der Wiel (2010) provides evidence referring to a policy reform of terms of notice in the Netherlands. A further connection between EPL and wage inequality can be found in comparative analysis: Bryson et al. (2012) show that higher labor (and product) market regulation is associated with lower use of incentive pay (ranging from 10% of covered workers in Portugal to 50% of the workforce in the United States). Inasmuch as incentive-pay schemes increase within-group earnings inequality (Lemieux et al., 2009), this induces a negative correlation at the aggregate level between earnings inequality and EPL indexes.

18.4.6.6 Labor Market Policies

Kluve (2010) provides an extensive meta-analysis based on a data set that comprises 137 active labor market program evaluations from 19 countries. Four main categories of ALMP are considered across European countries: (i) training programs, (ii) private-sector incentive schemes (such as wage subsidies to private firms and start-up grants), (iii) direct public employment programs, and (iv) "services and sanctions," a category comprising all measures aimed at increasing job search efficiency, such as counselling and monitoring, job-search assistance, and corresponding sanctions in case of noncompliance. His main finding is that traditional training programs have a modest significant positive impact on postprogram employment rates, but both private-sector incentive programs and services and sanctions show a significantly better performance. Evaluations of direct employment programs, on the other hand, are around 25% points less likely to estimate a significant positive impact on postprogram employment outcomes. Although effectiveness is here defined in terms of employment impact, they can be easily mapped one-to-one to wage inequality whenever the unemployed are taken into the picture.

[119] Similar results are found in Messina and Vallanti (2007). However results significantly differ when using aggregate or microdata. For example, using a German employer–employee matched data set. Bauer et al. (2007) do not find any evidence of variable enforcement of dismissal protection legislation on the employment dynamics in small establishments. Considering that labor churning is typically associated to increased earnings variability, their result would imply lack of correlation between employment protection and wage inequality. Analogous lack of significant impact of firing restrictions is found by in Martins (2009).

18.4.6.7 Stepwise Institutional Change

A new and different line of argument regarding institutions is nicely summarized by Boeri (2011). After reviewing existing institutional differences among European countries and stressing their persistence over time, he proposes a taxonomy of institutional changes (*reforms*), in terms of orientation and phasing-in. The orientation concerns the question whether they reduce (e.g., by making employment protection less strict and/or unemployment benefits less generous or by expanding the scope of activation programs) or increase the wedge (e.g., by increasing labor-supply-reducing taxes on relatively low-paid jobs) introduced by LMIs between supply and demand. Boeri accordingly classifies a reform as either decreasing or increasing the (institutional) wedge. The second characteristic relates to the phasing-in of reforms: this can be either complete or partial. In the former case, the change in the regulation eventually involves everybody. In the latter case, even at the steady state, the reform is confined to a subset of the population. The timing is also important. Even a complete phasing-in may involve a very long transitional period, so that the steady-state institutional configuration is attained beyond the planning horizon of management's potential involvement by the reform (Boeri, 2011, 1184). A two-tier reform is then defined as the case involving either a partial phasing-in or when its complete phasing-in requires more than 30 years, the average length of the working life in many countries. According to data collected over the period 1980–2007 for the European Union, the two-tier pattern is prevailing in most of the institutional dimensions. This has obvious implications in terms of earnings inequality, especially between insiders and new entrants (typically women and youngsters). With the help of a search model à la Pissarides-Mortensen, Boeri shows that institutions affect the threshold below which it is no longer convenient for either the employer or the employee to continue the work relationship. Even if the underlying inequality pattern depends on idiosyncratic shocks hitting individual productivity, the boundaries of the distribution of realized wages are institutionally determined, owing to variation in the equilibrium unemployment. According to the model an *increase in unemployment benefits* raises the reservation productivity at which matches are dissolved as the outside option of workers has improved: in equilibrium there is a higher probability of job loss, a lower job finding rate, higher unemployment and average wage.[120] Conversely an *increase in firing taxes* has the opposite effect of maintaining alive jobs with a lower match productivity. This reduces the gross job destruction rate and positively affects wages. An *increase in employment conditional incentives* (modeled as an employment subsidy) makes the labor market tighter, and increases the duration of jobs at the expenses of a decline in entry wages. Finally, an *increase in the activation scheme* reducing recruitment costs features higher

[120] For simplicity Boeri assumes that any unemployed person is entitled to the benefit, but actually this depends on the length of the contribution period and/or on belonging to specific categories (married/unmarried, with/without children, sector of employment, age).

job finding and job-loss rates, whereas the effects on unemployment and the average wage are ambiguous. When liberalizing (wedge-reducing) reforms are applied to only a fraction of workers (temporarily creating a dual labor market), then earnings inequality expands: insiders enjoy a surplus over outsiders at the same productivity levels, which is increasing in the difference in replacement rate offered to the unemployed (coming from long-tenured jobs with respect to those coming from short-tenured jobs), in the employment conditional incentive and in firing taxes, which matter more when workers have more bargaining power.

Returning to the more general, internationally comparative literature, developed by Blau and Kahn (2002) and others, we find the contribution of Koeniger et al. (2007) who look beyond cross-sectional differences at the comparative evolution of wage inequality over time, and extend to more OECD countries over a longer period, focused on overall wage inequality of males taken from the OECD database. They treat the various institutions (union density, union coordination/centralization, the minimum wage, employment protection, unemployment benefit generosity and duration, and the tax wedge) simultaneously and also model some interactions. On the demand side they control for the aggregate economy (unemployment rate), the relative supply of skills, international trade (import intensity), and technology (R&D intensity). They add some counterfactual simulations, including one that attributes US institutions to the other countries. They find compressing effects on the wage distribution of most institutions which explain at least as much as trade and technology do on the demand side. Applying American regulations would increase wage inequality in Continental Europe by 50–80%. The authors observe, however, that endogenizing the institutions, that means accounting for their dependence on supply and demand, will likely reduce the effects somewhat.

Finally, as we have observed above, the context of household (joint) labor supply potentially augments the number of institutions that need to be addressed, adding parental leave, maternity leave, part-time work regulations, and any other institution affecting the flexible use of working hours. Analyses of this (e.g., Dupuy and Fernández-Kranz, 2011; Thévenon and Solaz, 2013) are few, and they are focused on employment chances and/or pay penalties of gender/motherhood/family, not on the wage dispersion.

18.4.7 Summing Up

Over time the literature seems to have gone in two different directions that tend to grow further apart—not in the sense of interactions (one retorting to the other) but in the sense of integrating the approaches into one framework. Freeman (2007, p. 24) signaled the risk of creating the social science equivalence of "epicycles"—aimed at preserving Ptolemaic views on the earth as the center of the universe—for the institutional approach. However, the same danger may be looming for the supply-and-demand approach, which has been adding tasks, offshoring, and consumer preferences, in an attempt to dispel doubts about the relative demand of skills as a tautology. The institutional approach faces an abundance of institutions for which it lacks a clear criterion of choice; the supply-and-demand approach by contrast is challenged by the need for finding better empirical measures of technological change. However, a fortunate effect of the interactions just mentioned has been the great interest that is now taken in the very data on wage inequality. The take on the data's properties, advantages, and disadvantages has greatly improved over time. Consideration of the data at later points in time alter the stylized facts and also show that consensus explanations may be temporary and can break down when data for later periods become available and shine a different light on preceding periods. In spite of this, the prime aim of future work on both sides should be to integrate the other side into the framework. Pursuing that may be more a problem of empirical method for the institutional side, and on the demand-and-supply side the problem may be more on the theoretical side as long as institutions continue to be viewed as alien bodies. For both sides there is a perspective of work to do at the firm level. Matched employer-employee data (Cardoso, 2010; Lane, 2009) can help enlighten the role of both institutions and labor supply and demand (see, e.g., Andersson et al., 2006; Matano and Natichioni, 2011 for some interesting attempts). In addition, though much attention has been paid to data quality, a better grasp of the customary use of inequality measures seems desirable.

18.5. LMIs AND WAGE INEQUALITY: AN EMPIRICAL ASSESSMENT

In this section we present an accounting framework and an empirical model aiming to assess the contribution of LMIs to shaping earnings inequality. Here we face the problem of identifying who are benefiting from (or disadvantaged by) the action of a specific LMI. Before we have mentioned the stepwise changes introduced by many institutional reforms, which seem to create two-tier systems (Boeri, 2011), implying that the effect of institutions on earnings inequality may significantly differ across age cohorts. To deal with this, the ideal data set would be longitudinal, in order to be able to compute inequality measures over the lifetime of earnings, conditional of attrition in the sample creation. In addition, measuring institutions is not an easy task. Even if we restrict ourselves to the notion of institutions as rules inducing deviations from competitive market equilibria in economic transactions, these rules are still difficult to measure, because they often treat individuals differently or affect their behavior differently (think, for example, of taxes and benefits, which are almost always conditional to family composition—Boeri and Van Ours, 2008). Rules and norms change rather smoothly over time; in the definition used by Boeri (2011), reforms are rarely radical, and therefore it can take a significant amount of time before a minimum detectable effect may be observable. Despite these limitations, a significant literature has studied the correlation between institutional measures and earnings inequality measures (Alderson and Nielsen, 2002; Rueda and Pontusson, 2000; Wallerstein, 1999—more recently Kierzenkowski and Koske, 2012; Scheve and Stasavage, 2009). It exploits, in turn, cross-country and/or over-time variations of the institutions to arrive at estimates of the correlation with earnings inequality. In many instances, the dependent variable (the inequality measures) are derived from secondary sources, and do not always allow for measures that are fully comparable across countries (Atkinson and Brandolini, 2001). Some studies have computed their own inequality measures, relying on existing projects of data harmonization across countries (Atkinson, 2007a,b; Checchi and Garcia Peñalosa, 2008). We have followed here the same line of research, by computing appropriate indices of earnings inequality from SILC and PSID data sets, described in Section 18.3. Given the absence of natural experiments to obtain estimates of the causal impact of specific norms onto the relevant inequality measures, we will obtain at best correlations between institutional measures and inequalities. In Section 18.5.1, we consider a simple accounting scheme in order to discuss the correlation of market equilibria, institutions, and between-group inequality, whereas in Section 18.5.2 we provide a decomposition of the within-group earnings inequality and correlate these measures with proxies for institutions. In Section 18.5.3 we correlate inequality measured across age cohorts with past institutional measures, finding evidence of inequality-reducing impact of unions and minimum wages. Section 18.5.4 discusses the results.

A simple accounting scheme is plotted in Figure 18.24, which adopts the core of a scheme presented in OECD (2011) and elaborates on that. It describes the process of generating earnings inequality in an institutional framework. Starting components, individual wages, and hours worked are clearly affected by either the bargaining activity of unions (where/when present and active) and/or by existing regulations (minimum wage, regulation on worked hours). This determines individual labor earnings among the employees, but the total level of employment (and its split between dependent and self-employment) are conditioned by existing taxation as well as by employment protection (because so-called self-employment may disguise dependent employment conditions, especially in the case of a single purchaser). In addition, the generosity of public benefits to those laid-off or unemployed also contributes to reducing earnings inequality in the bottom part of the distribution. Although we will not proceed further with our analysis in that direction, one should keep in mind that the list of potential institutions affecting earnings inequality at large should consider the household dimension. Half of the sample of the workforce population is concentrated in households where two members are employed (either as dependent or self-employed). As long as their earnings are not perfectly correlated, cohabitation (and expected income sharing) works as a shock absorber. However, one-fourth of the population does not possess this insurance, as they are single-person households who by definition lack such shielding from the unemployment risk.

18.5.1 A Simple Scheme to Account for Between-Group Inequality

To frame our theoretical expectations before moving to the econometrics, let us consider a simple model that considers a partition of the population into groups. As such, it may be

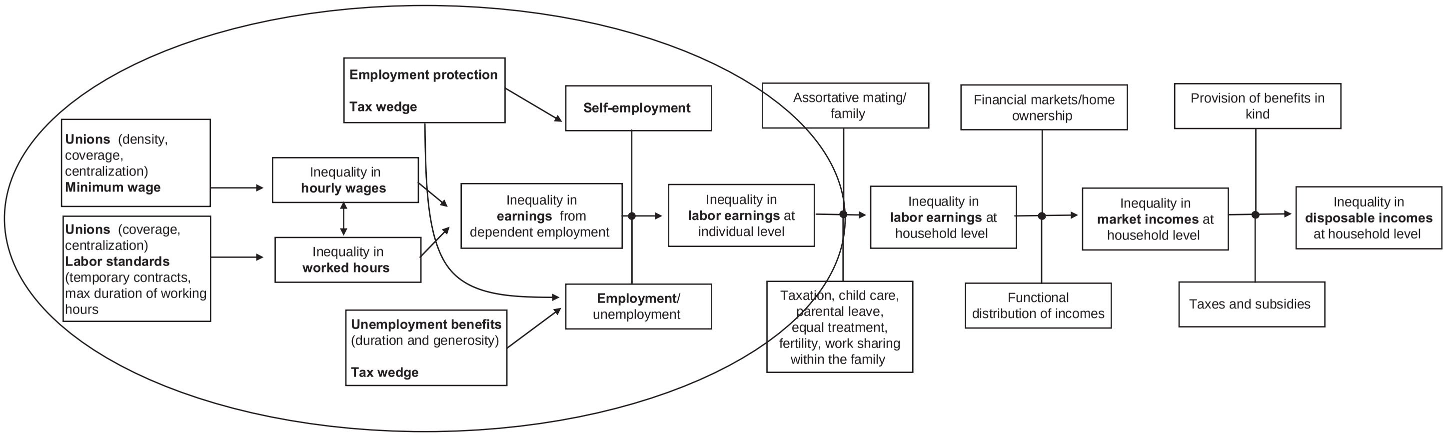


Figure 18.24 Accounting for the basic components of income inequality and the role of institutional measures. Source: *Adapted from OECD (2011, box 1).*

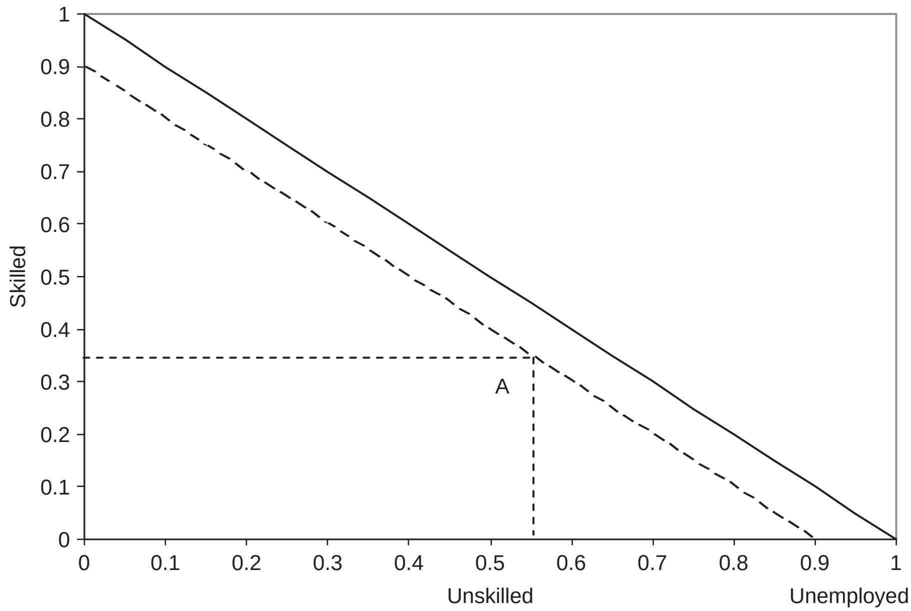


Figure 18.25 The distribution of the population.

considered appropriate to sketch the between-group component of inequality, whereas the between-component incorporates idiosyncratic components (including different marriage attitudes in each group), which are not necessarily connected to the institutional framework. This model builds on Atkinson and Bourguignon (2000) and Checchi and Garcia-Peñalosa (2008). If the workforce is composed by skilled and unskilled workers, a fraction of which may be unemployed, an inequality measure (Gini index) can be expressed (see Box 18.1)

$$\text{Gini}_{\text{earnings}} = f\left(\underset{\pm}{\alpha}, \underset{+}{\sigma}, \underset{\pm}{u}, \underset{-}{\gamma}\right)$$

where α indicates the share of skilled workers, σ wage differential between skilled and unskilled wage, u the unemployment rate, and γ the generosity of the unemployment benefit. This ideal population can be represented on the unitary simplex (see Figure 18.25), which has its empirical counterpart in our data set (see Figure 18.26). Although it is intuitive that earnings inequality is increasing in skill premium and decreasing in the generosity of the unemployment support scheme (conditional on the replacement rate being less than 100%), the effects of the other two parameters are ambiguous. Inequality is increasing in the skill composition as long as the initial fraction of skilled worker is small enough and/or not extremely well paid vis-à-vis the other unskilled

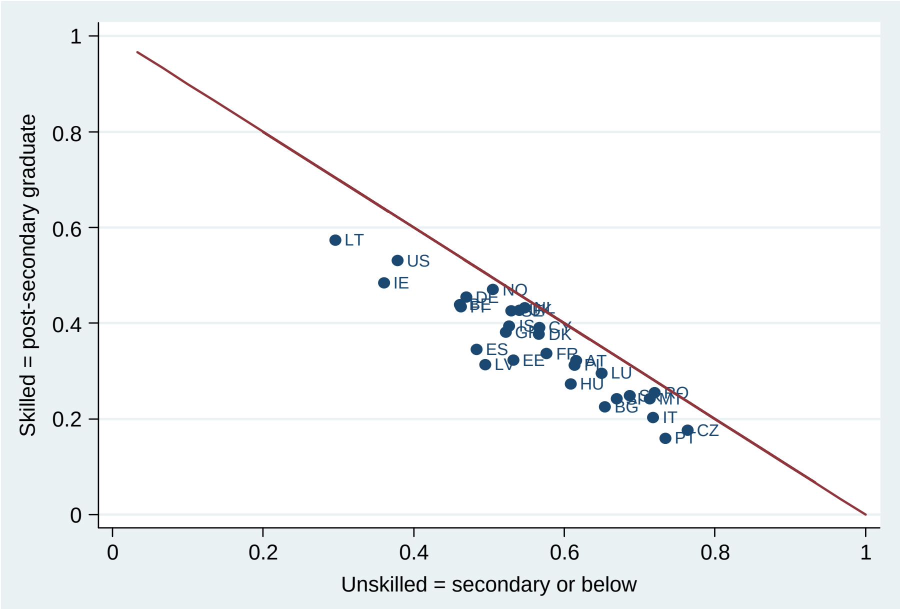


Figure 18.26 The distribution of the employee workforce (aged 20–55)—SILC 2010 and PSID 2011.

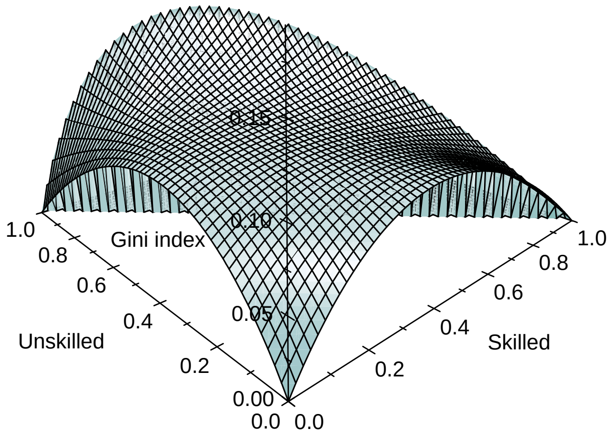


Figure 18.27 Plot of the Gini surface ($\gamma = 0.5$, $\sigma = 2$).

workers (i.e., the skill premium is small).[121] Eventually earnings inequality is increasing in unemployment rate in an intermediate range, while it exhibits negative correlation for high or low values (Figure 18.27).

[121] The ambiguous effect of α on Gini is not surprising because a change in α leads to Lorenz curves which cross each other, meaning that the change in the Gini will depend on how they cross each other; as a consequence other inequality measures may yield results in contradiction with the Gini index.

We are now in the position to discuss the relationship between earnings (between-groups) earnings inequality, market determinants, and LMIs. Among the four parameters identified by the model, one is partly independent from LMI. The skill composition of the employed (parameter α) depends on the interplay between demand and supply of skills. Demand for skill may be related to the technological development of an economy, which, in turns, relates to the international distribution of production and the possibility of off-shoring (Acemoglu and Autor, 2011, 2012). The supply of skills is the output of the educational system of a country, combined with expectations regarding wage premia. If we extend the notion of institutions to include educational systems, then this is the first determinant of wage inequality, which is nonlinearly related to earnings inequality (Leuven et al., 2004). Given intergenerational persistence in educational choice, the skill composition of the labor force changes rather smoothly across generations, and can be taken as given, at least in the short run.

By contrast, the return to skill (parameter σ) is jointly affected by competitive market forces and by institutions. In a competitive environment, this relative wage should be negatively correlated with the relative supply, as is slightly the case in Figure 18.28 (Katz and Autor, 1999). However there are significant deviations from such a relationship, which, among other factors, depend on the bargaining activity of unions (typically pursuing an egalitarian stance, aiming to tie wages to jobs and not to people—Visser and

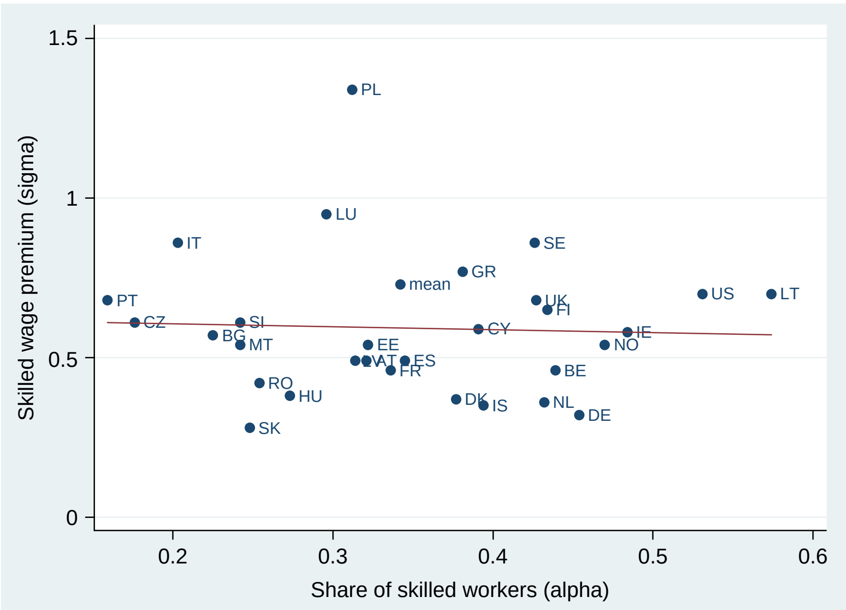


Figure 18.28 Return to skills and skill availability for dependent employees (aged 20–55)—SILC 2010 and PSID 2011.

Checchi, 2009; see also the role of wage scales described by Oliver, 2008) as well as the presence and coverage of minimum-wage legislation.

The unemployment benefit (parameter γ) has an uncontroversial effect of reducing earnings inequality when unemployed people are counted in. However, there is a general consensus that it has also a detrimental effect on the incentive to work, thereby raising the unemployment rate. Because the unemployment benefit can be thought of as a proxy for the outside option in wage bargaining or efficiency wage models, it also creates an upward wage push, which contributes to a positive correlation between benefit and unemployment. The overall effect is therefore $\frac{\partial \text{Gini}}{\partial \gamma} = \left.\frac{\partial \text{Gini}}{\partial \gamma}\right|_{u=\text{constant}} + \frac{\partial \text{Gini}}{\partial u} \cdot \frac{\partial u}{\partial \gamma}$ which can be either positive (for a high level of unemployment and/or a weak elasticity of unemployment to benefit) or negative (for a low level of unemployment and/or a high elasticity of unemployment to benefit). In our sample, the correlation tends to be positive (see Figure 18.29—however, this concerns short-run unemployment rates, whereas such a correlation should be studied using multiperiod unemployment rate in order to dispense with cyclical fluctuations). Once again, this is not the unique determinant of the unemployment rate (parameter u), because in a more general equilibrium model it depends on the state of the aggregate demand as well as on the average labor cost, which should incorporate the tax wedge. In addition, it may also be correlated with many other LMI variables, sometimes referred as determinants of the NAIRU (Nickell, 1997).

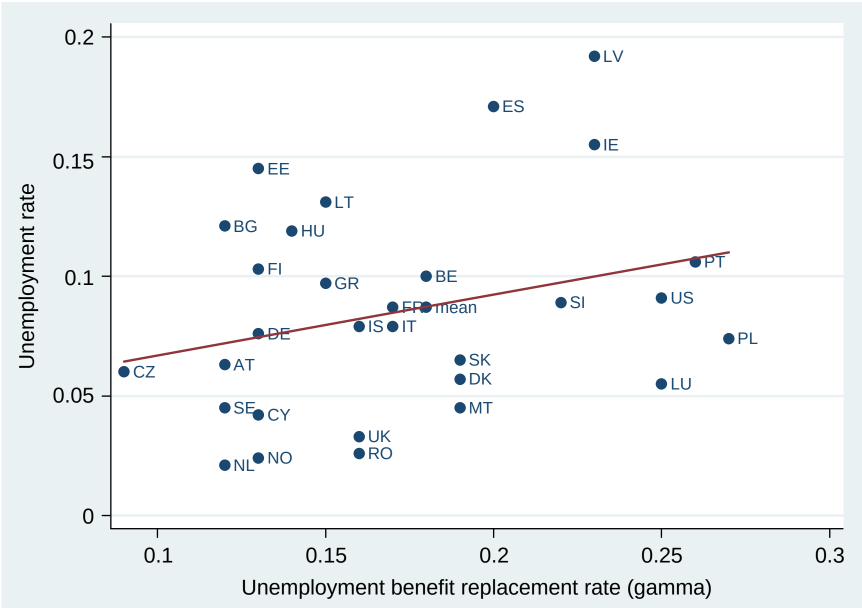


Figure 18.29 Unemployment benefit and unemployment rate—SILC 2010 and PSID 2011.

Still on the side of between-group inequality, we have purposely ignored the functional distribution of income between profit and wages, even though some of these parameters may be correlated to the labor share in the value added. Checchi and Peñalosa (2008) have shown that the same LMI affecting the functional distribution of value added, also affect the distribution of income sources at the individual level, thus modifying income inequality at the aggregate level.

Box 18.1 A model for between-group inequality in earnings

Let us suppose that the workforce has unitary measure and is composed of three groups of individuals:

(i) a fraction $\alpha \in (0,1)$ of the employed is made of skilled workers, earning a wage $w^s = (1+\sigma)w^u$, where $\sigma > 0$ is the skill premium[122];

(ii) a complementary fraction $(1-\alpha)$ is given by unskilled workers, who obtain a wage w^u.

(iii) a fraction u is unemployed and get a benefit $b = \gamma\overline{w}$ where $\gamma \in [0,1]$ is the replacement rate and $\overline{w}$ is the average wage within the employed labor force; skilled and unskilled workers experience the same unemployment rate.[123]

Each economy can be described by two coordinates, the unemployment rate u and the workforce composition α, and can be represented as a point in the unitary simplex. In Figure 18.25 the economy corresponding to point A is characterized by 10% of unemployment, two-thirds of unskilled employees and one-third of skilled ones. The same scheme could be applied to other dual partitioning of the labor force (young/old, male/female, native/foreign, etc.). The actual distribution of the population across different countries in our sample of analysis is reported in Figure 18.26.[124]

Our reference measure of inequality, the Gini concentration index, can provide a measure of the between-group inequality when computed in this simplified population by considering the subgroup differences, obtaining the following expression:

[122] We do not consider the presence of a fourth fraction of rich capitalists, as in Alvaredo (2011), who shows that when their population share is negligible (as in the case of top incomes), the Gini inequality index G_{incomes} can be approximated by $G^*_{\text{incomes}} \cdot (1-S) + S$, where G^*_{incomes} is the Gini coefficient for the rest of the population and S is the share of total income accruing to the rich fraction of the population. Similarly, the model could be complicated by introducing a third group of workers with an intermediate level of skills, to account for the possibility of polarization.

[123] This simplifies the analysis, avoiding to model relative labor demand, which would allow for modeling a differential impact of institutions on worker subgroups: "Any observer of European labour markets in the last 30 years of the twentieth century would agree that it is a good stylized description of these markets to think of the labour market for high-skill workers as in equilibrium, with wages that adjust to offset demand and supply imbalances, while the low-skill labour market is in disequilibrium, with involuntary unemployment and unresponsive real wages" (Saint-Paul, 2000, 5).

[124] A worker is arbitrarily classified as skilled when possessing a postsecondary school degree. This explains why formerly planned economies exhibit such wide variations in skill endowments.

$$\text{Gini}_{\text{earnings}} = \frac{(1-u)^2\alpha(1-\alpha)[w^s - w^u] + (1-u)\alpha u[w^s - b] + (1-u)(1-\alpha)u[w^u - b]}{2[(1-u)\overline{w} + ub]} \tag{18.1}$$

Using previous definitions, Equation (18.1) can be reexpressed as

$$\begin{aligned}\text{Gini}_{\text{earnings}} &= \frac{(1-u)^2\alpha(1-\alpha)\sigma + (1-u)u(1-\gamma)(1+\alpha\sigma)}{2[1-u(1-\gamma)](1+\alpha\sigma)} \\ &= \frac{(1-u)^2\dfrac{\alpha(1-\alpha)\sigma}{(1+\alpha\sigma)} + (1-u)u(1-\gamma)}{2[1-u(1-\gamma)]}\end{aligned} \tag{18.2}$$

Thus, the (between-groups) inequality in the earnings distribution is parameterized over four characteristics: the employment rate $(1-u)$, the labor force composition α, the skill premium σ, and the generosity of the unemployment benefit γ. It is easy to show that $\frac{\partial \text{Gini}}{\partial \gamma} < 0$ and $\frac{\partial \text{Gini}}{\partial \sigma} > 0$, namely that other things constant, earnings inequality is increasing in skill premium and decreasing in the generosity of the unemployment support scheme. Less clear-cut results obtain with respect to the other two parameters. It can be proved that $\text{sign}\left[\frac{\partial \text{Gini}}{\partial \alpha}\right] = \text{sign}[1-\alpha(\alpha\sigma+2)]$, which is positive for $0 \leq \alpha < \frac{\sqrt{1+\sigma}-1}{\sigma}$. Thus, inequality is increasing in the skill composition as long as the initial fraction of skilled worker is small enough and/or not extremely well paid vis-à-vis the other unskilled workers (i.e., the skill premium σ is small). In the case of unemployment tedious calculations[125] prove that

$$\frac{\partial \text{Gini}}{\partial u} > 0 \quad \text{iff} \quad \frac{1}{1-\gamma}\left(1 - \sqrt{\gamma\frac{A\gamma-(1-\gamma)}{A-(1-\gamma)}}\right) < u < \frac{1}{1-\gamma}\left(1 + \sqrt{\gamma\frac{A\gamma-(1-\gamma)}{A-(1-\gamma)}}\right),$$

$$A = \frac{\alpha(1-\alpha)\sigma}{(1+\alpha\sigma)} < 1$$

Thus, earnings inequality is increasing in unemployment rate in an intermediate range, while it has a negative correlation for high or low values. The Gini surface over the unitary simplex is represented in Figure 18.27: notice that the hump-shape is consistent with the just-mentioned derivative.

So far we have only considered the between-group inequality, ignoring the within-group component, because the former is easier to correlate with LMIs. If we want to take into account both components in an explicit way, we need to resort to a decomposable inequality index, like the generalized entropy index (with $\alpha = 0$), known as mean logarithmic deviation $MLD = \frac{1}{n}\sum_{i=1}^{n} \lg\left(\frac{\overline{y}}{y_i}\right)$ (Jenkins, 1995). In the framework of the present model it can be decomposed as

[125] If we rewrite the Gini index as $\text{Gini} = \frac{(1-u)^2A + (1-u)uB}{2(1-uB)}$ where $A = \frac{\alpha(1-\alpha)\sigma}{1+\alpha\sigma} < 1$ and $B = (1-\gamma) < 1$, then $\text{sign}\left[\frac{\partial \text{Gini}}{\partial u}\right] = \text{sign}[-B(A-B)u^2 + 2(A-B)u + B - A(2-B)]$, which has two real roots under the sufficiency condition that $A > \frac{B}{1-B}$. These roots are given by $u_{1,2} = \frac{(A-B) \pm \sqrt{(A-B)(1-B)(A-B-AB)}}{B(A-B)} = \frac{1}{B}\left(1 \pm \sqrt{\frac{(1-B)(A(1-B)-B)}{(A-B)}}\right)$, which corresponds to what is reported in the text.

$$\mathrm{MLD_{earnings}} = \underbrace{\alpha(1-u)\cdot \mathrm{MLD_{skilled}} + (1-\alpha)(1-u)\cdot \mathrm{MLD_{unskilled}} + u\cdot \mathrm{MLD_{unemployed}}}_{\text{within-group inequality}}$$
$$\underbrace{+\,\alpha(1-u)\cdot \lg\left(\frac{\mu}{w^s}\right) + (1-\alpha)(1-u)\cdot \lg\left(\frac{\mu}{w^u}\right) + u\cdot \lg\left(\frac{\mu}{\gamma\overline{w}}\right)}_{\text{between-group inequality}} \tag{18.3}$$

where μ is the mean income in the population. So far we have neglected the funding of the unemployment benefit scheme (which could derive from profit and rent taxation). In such a case

$$\mu = (1-u)\overline{w} + u\gamma\overline{w} = (1-u(1-\gamma))\overline{w} = (1-u(1-\gamma))(\alpha w^s + (1-\alpha)w^u)$$
$$= (1-u(1-\gamma))(1+\alpha\sigma)w^u$$

On the contrary, if we impose a balanced budget, such that unemployment benefits are to be financed by earnings taxation, we require that $(1-u)t\overline{w} = u\gamma\overline{w}$, where t is the average tax rate. As a consequence

$$\mu = (1-u)\overline{w} + (1-u)t\overline{w} = (1-u)(1+t)(1+\alpha\sigma)w^u \tag{18.4}$$

If we replace definition (18.4) into Equation (18.3) we obtain

$$\mathrm{MLD_{earnings}} = \underbrace{\alpha(1-u)\cdot \mathrm{MLD_{skilled}} + (1-\alpha)(1-u)\cdot \mathrm{MLD_{unskilled}} + u\cdot \mathrm{MLD_{unemployed}}}_{\text{within-group inequality}}$$
$$\underbrace{+\,\alpha(1-u)\cdot \lg\left(\frac{(1-u)(1+t)(1+\alpha\sigma)}{1+\sigma}\right) + (1-\alpha)(1-u)\cdot \lg((1-u)(1+t)(1+\alpha\sigma)) + u\cdot \lg\left(\frac{(1+t)u}{t}\right)}_{\text{between-group inequality}} \tag{18.5}$$

It is easy to prove that the between-group component of MLD is increasing in σ and decreasing in γ (under the balanced-budget constraint). In addition, the between-group component is increasing in α for low values, but it changes sign above the threshold defined by $\alpha^* = \frac{1}{\lg(1+\sigma)} - \frac{1}{\sigma}$. The main difference with the Gini measure of inequality is that the gradient of the between-component with respect to the unemployment rate u takes the sign of $\left[1 - \frac{(1+\alpha\sigma)\gamma}{(1+\sigma^\alpha)}\right]$ suggesting that inequality is increasing whenever the replacement rate and/or the wage premium are low.

If we are to check the predictive ability of this simple model, we can use observed sample parameters $(\alpha, \sigma, u, \gamma)$ to predict earnings inequality in each country, well aware that this captures only the between-group component. We define as skilled workers all employees holding a postsecondary degree, and compute the skilled wage as their mean wage. Correspondingly, we define as unskilled all the remaining employees (and obtain their wage); finally, we compute the unemployment share and their mean benefit. The relevant parameters, which are needed for the between-group inequality measures, are reported in Table 18.A4. In column 10, we report the estimated Gini, which has to be compared

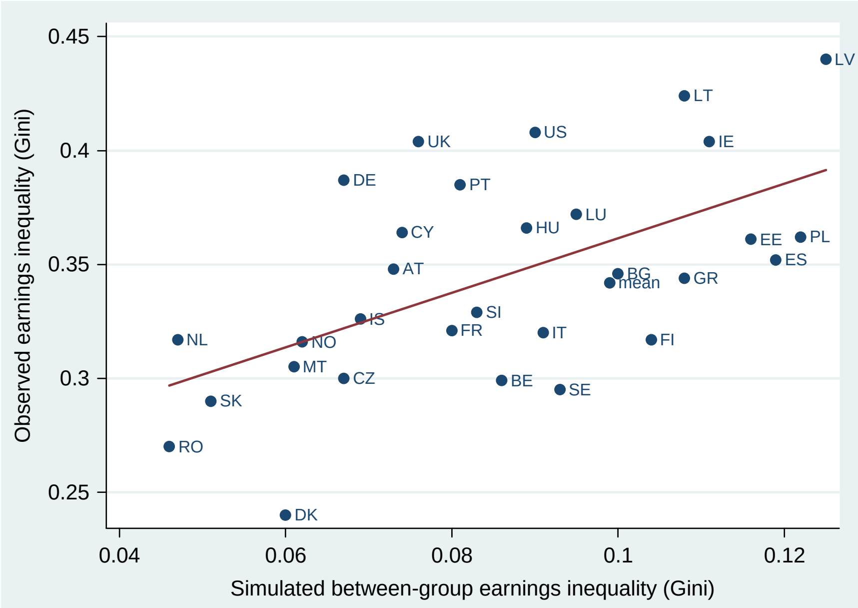


Figure 18.30 The between-group component of earnings inequality—SILC 2010 and PSID 2011.

with the actual one computed on the same data set in column 11. The two coefficients are highly correlated (rank correlation coefficient is 0.57).[126]

Using the Gini index computed over four parameters, we can claim that the between-group component accounts for almost one-third of overall earnings inequality, the remainder being attributable to individual heterogeneity (age, gender, finer partition of educational attainments—including variations of hours). It is rather surprising that such a simple model, based on four parameters only, is able to account for a significant portion of the observed cross-country differences in earnings inequality. Looking at Figure 18.30 we notice that some countries (lying to the right of the regression line) are characterized by higher-than-the-mean between-group inequality (or lower-than-the-mean overall earnings inequality): not surprisingly the Nordic and the Mediterranean countries (except Portugal) are on this side, indicating that in these countries institutions may help to reduce the corresponding within-group inequality. On the left side of the regression line, however, we find the liberal market economies (United States, United Kingdom, and Ireland) and some transition economies (Latvia, Lithuania, and Hungary) as well as some

[126] Regressing the observed Gini in labor earnings onto the simulated one computed according to Equation (18.2) yields the following estimation: $\text{Gini}_{\text{observed}} = \underset{(0.03)}{0.24} + \underset{(0.32)}{1.20} \cdot \text{Gini}_{\text{simulated}}$ with an $R^2 = 0.33$.

continental European country (such as Germany and the Netherlands). These countries are characterized by individual rather than collective wage setting, thus raising the between-group component of earnings inequality.

18.5.2 The Within-Group Inequality and the Role of LMIs

We now consider the within-group component of inequality. To obtain an exact decomposition of earnings inequality for employees, we abstract from self-employment (we think it is potentially affected by existing labor market regulations, but it often also records negative incomes that are not easily dealt by inequality measures), and we restrict ourselves to individuals aged below 55 (to minimize country differences attributable to a different extent of early retirement[127]) who receive either a positive income from dependent employment or from unemployment benefit. Using the mean log deviation to decompose earnings inequality, we find that on average the between-component accounts for one-fifth of the observed inequality, being highest in Portugal (30%), Hungary (28%), and Slovenia (28%) and lowest in Sweden (7%), Norway (8%), and the Netherlands (11%) (see Table 18.3).

The within-group component follows common patterns: inequality is highest among the unemployed,[128] but its contribution to the within-group component is limited, the country average being 16%. Skilled workers are characterized by higher earnings inequality than the unskilled ones, and this is not surprising once we consider that their wage will more frequently be determined by individual bargaining. The unskilled workers (who on average comprise 57% of the workforce) do contribute half of total within-group inequality, and it is here that we may expect to find the strongest impact of LMIs (especially the minimum wage and bargaining activity of unions).[129]

[127] The SILC codebook allows for the classification as unemployed of early retired workers if they perceive themselves as such ("Early retirement for economic reasons can be included here according to the respondent's feeling, i.e., a person in early retirement for economic reasons will be included here if he/she classifies him/herself as unemployed" (Eurostat, Description of Target Variables: Cross-sectional and Longitudinal 2010 operation (Version February 2010, 139).

[128] Inequality among (unemployment) benefit recipients is significantly affected by the duration of unemployment spells, by differences in the entitlement rights and in the take-up rates. Although we do not have adequate data to cope with all these factors, if we just replace the current figures for the benefit with its monthly average (simply dividing the yearly received subsidy by months in unemployment) we obtain that the inequality in monthly unemployment benefit significantly declines for some countries (Austria, Czech Republic, Netherlands, Norway) but it increases in others (Estonia, Ireland, Italy), the country average of MLD remains almost unchanged (from 0.388 to 0.381).

[129] Freeman and Schettkat (2001) follow a similar approach when comparing US and German earnings inequality, showing that inequality within each educational group is higher in the former country, and they attribute it to the role of bargaining structures.

Table 18.3 Earnings inequality decomposition—dependent employees or unemployed (mean log deviation)—SILC 2010 and PSID 2011

	Decomposition of total inequality			Decomposition of the within-group component					
	Overall yearly gross earnings inequality	**Between group inequality**	**Within group inequality**	**Population share of skilled workers**	**Inequality in yearly gross earnings of skilled workers**	**Population share of unskilled workers**	**Inequality in yearly gross earnings of unskilled workers**	**Population share of unemployed workers (receiving a positive benefit)**	**Inequality in unemployment benefits (conditional on being unemployed)**
Austria	0.234	0.031	0.203	0.321	0.223	0.616	0.181	0.063	0.327
Belgium	0.179	0.048	0.131	0.439	0.124	0.461	0.119	0.100	0.214
Bulgaria	0.241	0.046	0.195	0.225	0.176	0.654	0.161	0.121	0.410
Cyprus	0.263	0.031	0.232	0.391	0.255	0.567	0.214	0.042	0.270
Czech Republic	0.184	0.033	0.151	0.176	0.171	0.764	0.119	0.060	0.494
Denmark	0.129	0.016	0.113	0.376	0.111	0.566	0.101	0.057	0.243
Estonia	0.270	0.033	0.237	0.322	0.200	0.532	0.194	0.145	0.475
Finland	0.204	0.049	0.155	0.434	0.145	0.462	0.132	0.103	0.294
France	0.229	0.031	0.198	0.336	0.197	0.576	0.173	0.087	0.367
Germany	0.334	0.069	0.265	0.454	0.234	0.470	0.276	0.076	0.381
Greece	0.224	0.031	0.193	0.381	0.197	0.522	0.164	0.097	0.328
Hungary	0.250	0.071	0.179	0.273	0.206	0.608	0.150	0.119	0.262
Iceland	0.218	0.027	0.191	0.394	0.169	0.527	0.178	0.079	0.384
Ireland	0.316	0.048	0.268	0.484	0.250	0.361	0.233	0.155	0.406
Italy	0.224	0.027	0.197	0.203	0.205	0.718	0.169	0.079	0.436
Latvia	0.418	0.076	0.342	0.314	0.279	0.495	0.261	0.192	0.653
Lithuania	0.377	0.061	0.316	0.574	0.295	0.296	0.248	0.131	0.566
Luxembourg	0.260	0.060	0.200	0.296	0.205	0.649	0.196	0.055	0.213
Malta	0.199	0.029	0.170	0.242	0.179	0.713	0.150	0.045	0.453
Netherlands	0.200	0.023	0.177	0.432	0.172	0.547	0.167	0.021	0.563
Norway	0.230	0.018	0.212	0.470	0.204	0.505	0.200	0.024	0.605

Continued

Table 18.3 Earnings inequality decomposition—dependent employees or unemployed (mean log deviation)—SILC 2010 and PSID 2011—cont'd

	Decomposition of total inequality			Decomposition of the within-group component					
	Overall yearly gross earnings inequality	Between group inequality	Within group inequality	Population share of skilled workers	Inequality in yearly gross earnings of skilled workers	Population share of unskilled workers	Inequality in yearly gross earnings of unskilled workers	Population share of unemployed workers (receiving a positive benefit)	Inequality in unemployment benefits (conditional on being unemployed)
Poland	0.253	0.038	0.215	0.312	0.226	0.614	0.188	0.074	0.387
Portugal	0.259	0.078	0.181	0.159	0.246	0.734	0.163	0.106	0.204
Romania	0.121	0.032	0.089	0.254	0.111	0.720	0.078	0.026	0.176
Slovak Republic	0.180	0.029	0.151	0.248	0.172	0.687	0.113	0.065	0.476
Slovenia	0.229	0.064	0.165	0.242	0.189	0.669	0.116	0.089	0.471
Spain	0.249	0.057	0.192	0.345	0.177	0.483	0.171	0.171	0.284
Sweden	0.230	0.016	0.214	0.426	0.245	0.530	0.166	0.045	0.484
United Kingdom	0.306	0.058	0.248	0.427	0.261	0.541	0.231	0.033	0.359
United States	0.339	0.045	0.294	0.483	0.310	0.468	0.271	0.049	0.347
Average	0.245	0.044	0.201	0.340	0.203	0.575	0.174	0.085	0.388

If we now consider the potential role of LMI in shaping the wage distribution within workers' types, we do expect a differential impact according to the way in which different workers are affected.[130] We spent some effort to collecting consistent information on institutional variables for the same countries, mostly from various OECD data sets. We tried to build long series in order to match individuals of different age cohorts to the institutional setup prevailing either at the beginning of their work careers or during their entire career. Data sources and descriptive statistics are in Appendix C.

Table 18.4 summarizes our theoretical expectations, mostly deduced from the existing literature. Betcherman (2012) reviews the empirical literature on the correlation between different institutional dimensions and earnings inequality. He concludes that the minimum wage is the less contentious among the institutional impact, being associated to an improvement in the bottom tail of the wage distribution, at least for the formal sector. Neumark and Wascher (2008) do not contest the inequality-reducing impact of minimum wage (by creating a spike at the relevant threshold and/or inducing upward spillover effect across the entire wage distribution), though they stress the contemporaneous disemployment effect on low-wage earners, raising doubts about the overall effect on inequality at household level.[131]

The effect of unions is mixed, combining a reduction of within-group inequality (among formal dependent employment, especially in terms of skill premium—Koeniger et al., 2007) and a potential increase in the wage gap between union-covered sectors and nonunion-covered sectors (including informal employment). Using cross-country data, Visser and Checchi (2009) find that union presence is associated with lower within-group inequality, because both the gender gap and the return to education are negatively correlated with union density.[132,133] As a consequence, the skill premium declines, both as a result of wage compression and as a consequence of the incentives to over-invest in education. In addition, union presence is also associated to

[130] Eichhorst et al. (2008) provide a recent review of how LMIs are measured and their impacts on unemployment.

[131] Among the long-run impacts they also list the inhibiting impact on skill acquisition for youngsters, which will split over into greater earnings inequality in the future. Thus, they conclude, "Minimum wages do not deliver on their goal of improving the lives of low-wage workers, low-skill individuals, and low-income families" (p. 293).

[132] The egalitarian attitude of workers' unions has been rationalized by Agell and Lommerud (1992) using the argument that high-productivity risk-adverse workers may prefer pay compression in the absence of a market for private insurance.

[133] We do not consider here that institutions may operate in a complementary way, through interactions. In particular employment protection reinforces the impact of union density on unemployment and wage bargaining (Belot and van Ours, 2004). Fiori et al. (2012) provide an empirical application of Blanchard and Giavazzi (2003), which shows the substitutability of product and labor market reforms in terms of employment impact.

Table 18.4 Theoretical expectations of the effects of institutions on earnings inequality

Labor market institutions	Between groups	Within groups	Overall impact on earnings inequality
Minimum wage (measured by ratio to median wage)	• Raises the bottom tail of hourly wage, mostly for the unskilled	• Raise the bottom tail (typically populated by marginal workers)	• Reducing inequality in hourly wages—overall effects depend on hours dynamics
Union presence (measured by union density, coverage, centralization and/or coordination, strike activity)	• Compresses the skill premium (*equal pay for equal work*) • Expands the union wage gap (between union and nonunion sectors/jobs)	• Reduce inequality in hours (control/opposition to overtime, regulation of part-time, work sharing as alternative to layoffs) • Reduces gender wage gap, thus favoring work-sharing within the family and female participation	• Reducing (ambiguous when unemployment effects are taken into account)
Employment protection (measured by OECD summary index)	• Lowers unskilled wage when inducing people to retain unproductive jobs • Increases long-term unemployment	• Reduces job flows in/out of unemployment • Discourages labor market entry of marginal workers (young, women)	• Ambiguous
Unemployment benefit (measured by replacement rate and public expenditure in passive labor market policies)	• Raises the income of the unemployed • Raises the outside option, thus augmenting the bargaining power of unions • Lowers the incentive to job search	• Potential subsidy traps (especially on second earner, because the reservation wage is positively correlated with first earner)	• Ambiguous

Tax wedge (measured by the ratio between labor cost and take home pay)	• Increases unemployment (if the employer is unable to transfer the burden onto the employee)	• When altering labor cost (if they cannot be shifted to workers), taxes and payroll taxes alter the relative employment of worker subgroups • Even within the household, EITC (earned income tax credit) measures may favor joint participation of spouses to the labor market, especially when part time is easily available	• Increases personal earnings inequality (because presence of part-timers) but may reduce household earnings inequality (because presence of an additional income in the household)
Active labor market policies (measured by public expenditure on GDP) Child/old people care facilities (availability of ECCE facilities, parental leave)	• Reduces unemployment	• Increasing labor market participation, possibly with reduced hours • Increasing female participation, it brings in additional workers into employment	• Ambiguous (due to the combined effect of participation and hours) • Ambiguous (due to a compositional effect)

unemployment, though correlation may go in different directions: union density seems associated with higher unemployment (Bertola et al., 2007; Flaig and Rottmann, 2011; Nickell et al., 2005), and centralized bargaining seems to attenuate this negative effect (Bassanini and Duval, 2006; Nickell, 1997—see also Glyn et al. (2003) for a critical review of these results). Thus, the overall effect of unions on earnings inequality remains uncertain.

The results of employment protection legislation are less clear-cut. OECD (2011, 2012) show that EPL and wage coordination have a negative effect on earning inequality, while tax wage and wage coverage have a positive effect. The proposed rationalization is that unskilled workers are favored by firing restriction, raising their relative bargaining power relative to skilled ones.[134]

Unemployment benefits, active labor market policies, and the tax wedge may play an indirect role, via the impact on aggregate employment (or unemployment). The tax wedge in particular has been found to be significantly and positively correlated to the unemployment rate (Flaig and Rottmann, 2011; Nickell et al., 2005).[135] But these two institutions also affect different groups of workers in different ways, especially along the gender divide (Bertola et al., 2007): as a consequence, they may impact on the household distribution of earnings via changes in the redistribution of work opportunities within the family. In addition, when aiming to decompose the contribution to inequality associated with hourly wages and hours worked, the legal framework (limitation to part-time, family, or individual taxation) may lead to opposite impacts on labor supply, the corresponding employment and wage outcomes. Possibly for these reasons, we have not found consensus on this dimension in the literature, and therefore we will let the data speak.

Work redistribution within the household may also be affected by parental leave opportunities and child care provisions (Thévenon and Solaz, 2013). As long as these institutional dimensions favor female participation, they should reduce earnings inequality measured at the household level, but they may increase inequality at the individual level, owing to a larger fraction of part-timers in the economy. However, these results are conditional on parental leave not exceeding a specific threshold, because otherwise it may produce a reduction in labor supply.[136] In addition, as long-mandated parental leave

[134] A similar argument can be found in Koeniger et al. (2007), where employment protection has stronger effect for less-qualified workers.

[135] Flaig and Rottmann (2011, 19) conclude from their cross-country analysis covering 19 OECD countries over the 1960–2000 period that "[a] tighter employment protection legislation, a more generous unemployment insurance system and a higher tax burden of labour income increase the medium term development of the unemployment rate, whereas a higher centralization of the wage bargaining process lowers unemployment. Union density has no clear effect and seems to be unimportant."

[136] Lalive et al. (2011) study the complementarity between job protection associated with parental leave and financial support to new parents, showing that either policy instrument has a detrimental effect on female labor supply in the medium run.

may raise female supply in the labor market, it may also exert a downward pressure on their relative wage, thus contributing to increased inequality (which, however, is not found in the limited data analyzed by Thévenon and Solaz, 2013). Also in such a case, we may let the data speak.

A serious problem in assessing the impact of single institutions on labor market outcomes is that some institutions are likely to interact with each other, in a positive or in a negative way. Consider, for example, the role of workers' unions, which is typically correlated in a negative way to earnings inequality. The presence of unions is strengthened by employment protection legislation, but is weakened by the presence of minimum-wage provisions.[137] Similarly, the tax wedge may have a significant impact on employment in a country where the (after-tax) minimum wage is relatively high because part of the wedge will be passed on to wages at a higher level. In some countries (such as France and Belgium) rebates on payroll taxes for low-wage workers significantly impact on their employability.

Addressing the issue of institutional complementarity opens up another set of literature, which is typically analyzed by political economy (Amable, 2003; Hall and Soskice, 2001). From an empirical point of view it does require a sufficient number of degrees of freedom (either in terms of variety of countries or in terms of repeated observations over the same country). Just as descriptive evidence, the sample bivariate correlations between the inequality measures presented in Table 18.4 and the LMIs described in Appendix C are presented in Table 18.5.[138] Exploiting the decomposability of the Mean Log Deviation, we have considered six dimensions of earnings inequality: its overall measure, the decomposition into between-group and within-group, and the contributions to the within-component attributable to each group of workers (skilled, unskilled, and unemployed).[139]

They confirm that union presence (either measured by union density or by coverage) may contribute to reducing earnings inequality, though in a different way. Union density seems statistically correlated with the between-group component, whereas the coverage of collective agreements (which assures equivalent treatment of all workers) exhibits a negative correlation with the within-component. Similar negative correlations are exhibited by employment protection with respect to the skilled worker group; analogously, parental leave facilities are negatively correlated to skilled wage inequality. It is

[137] Checchi and Lucifora (2002) discuss the complementarity/substitutability of LMIs with respect to union density.

[138] A review of existing data sets on LMIs is in Ochel (2005) and Eichhorst et al. (2008).

[139] By considering the contribution to inequality attributable to workers' groups we are combining two sources of variation: the group size and its internal inequality. Although the fraction of unemployed workers may be directly correlated to LMIs (such as unions or unemployment benefit), the skill composition of the labor force may be correlated with the quality and quantity of education available in the country in earlier decades.

Table 18.5 Correlation between labor market institutions (averages 2001–2010) and different component of earnings inequality (MLD)—SILC 2010 and PSID 2011

	Overall yearly gross earnings inequality	Between group inequality	Within group inequality	Inequality in yearly gross earnings attributable to skilled workers	Inequality in yearly gross earnings attributable to unskilled workers	Inequality in unemployment benefits attributable to unemployed
Union density	−0.415*	−0.420*	−0.346*	−0.112	−0.308*	−0.346*
Agreements coverage	−0.601*	−0.3512*	−0.584*	−0.391*	−0.378*	−0.436*
Centralization	−0.099	−0.223	−0.044	−0.036	−0.141	0.083
Strike activity	−0.280	−0.190	−0.268	−0.115	−0.343	−0.144
Minimum wage (Kaitz index)	0.127	0.3219*	0.043	0.026	−0.161	0.216
Employment protection legislation	−0.330	0.074	−0.410*	−0.626*	0.018	0.011
Unemployment benefit	0.125	0.010	0.142	0.061	−0.092	0.327*
Tax wedge	−0.241	−0.023	−0.273	−0.099	−0.316*	−0.188
Social expenditure	−0.229	−0.175	−0.190	0.035	−0.238	−0.267
Child care	−0.179	−0.040	−0.189	−0.252	0.037	−0.109
Parental leave	−0.320	−0.138	−0.318	−0.435*	−0.060	0.040
Tax treatment of household incomes	0.052	−0.170	0.120	0.292	−0.024	−0.117
Active labor market policies	−0.296	−0.222	−0.273	−0.047	−0.322*	−0.255
Passive labor market policies	−0.248	−0.068	−0.266	−0.089	−0.320*	−0.182

30 countries—*significant at 10%.

interesting to note that the generosity of the unemployment benefit seems to contribute positively to the inequality component attributable to the unemployed (even if we are unable to distinguish whether this is due to an increase of the unemployment rate or to a different distribution within the group). Not surprisingly, household or individual taxation does not affect wage inequality, because it may be ineffective in modifying household labor supply (Dingeldey, 2001). Active and passive labor market policies seem mostly effective in reducing the within-component of the earnings inequality of unskilled workers.

Overall these results are not satisfying in terms of statistical significance, suggesting that isolating a single institution at a specific point in time (even though here we are considering a decennial average) may not be the best strategy to investigate the association between inequality and institutions. Though it may sometimes be inevitable for empirical reasons, it does seem advisable to consider the degree of embeddedness of individual institutions in a collection of institutions to see whether one can lay more weight on analytical results obtained for one institution compared to another. For example, the strong legal nature of an institution may enhance its standalone effect. In addition, bivariate correlations are sensitive to the criticism of spurious correlation and also to omitted-variable bias. For this reason we now consider more robust methods to study the impact of institutions on earnings inequality.

18.5.3 Empirical Assessment

18.5.3.1 Cross-Sectional Approach

One crucial issue in the analysis of the role of LMI in shaping earnings inequality is the match of inequality computed from microdata to the corresponding institutional measures. If we correlate current inequality measured over workers of different ages (who therefore have been staying in the labor market for different durations) to the current union density (which is computed over the workers who are currently working) we are simply considering "industrial relations" regimes, without any claim of causality in one direction or the other. Such an exercise is conducted in Table 18.6, in which we consider three different dimensions of inequality (yearly earnings from dependent employment, hourly wages, and worked hours by dependent employees). In accordance with our previous between-group inequality decomposition (see Section 18.5.2), for each dimension we consider two market phenomena that are correlated with market forces: level of qualification of the labor force and level of employment (better captured by the female employment rate).[140] In all cases an increasing level of education in the

[140] Actually the skill level of the labor force is the joint outcome of the demand for education of the population and the institutional supply of schooling; however, replacing it with some measure of the strength of the institutional push toward education (such as the years of compulsory education) did not prove statistically significant.

Table 18.6 Gross earnings inequality (SILC 2010–PSID 2011) against market and labor market institutions (2001–2010) Effects—OLS

	1	2	3	4	5	6	7	8	9	10	11	12
Dependent variable	**Gini of gross annual employee earnings (working more than 1000 h per year)**			**Gini of gross hourly wages**			**Gini of employee (all durations)**			**Correlation of wages and hours**		
Population share with secondary degree	−0.309 [0.073]⋆	−0.06 [0.091]	−0.086 [0.131]	−0.424 [0.151]⋆⋆	−0.277 [0.190]	−0.367 [0.259]	−0.233 [0.083]⋆⋆	−0.139 [0.124]	−0.201 [0.148]	−0.095 [0.262]	−0.139 [0.124]	−0.201 [0.148]
Population share with postsecondary degree	−0.217 [0.059]⋆	−0.197 [0.067]⋆⋆	−0.211 [0.095]⋆⋆⋆	−0.235 [0.085]⋆⋆	−0.268 [0.100]⋆⋆	−0.289 [0.158]	−0.19 [0.078]⋆⋆	−0.189 [0.058]⋆	−0.189 [0.077]⋆⋆	0.083 [0.171]	−0.189 [0.058]⋆	−0.189 [0.077]⋆⋆
Female employment rate	−0.66 [0.211]⋆	−0.405 [0.175]⋆⋆	−0.528 [0.281]⋆⋆⋆	−0.816 [0.426]⋆⋆⋆	−0.679 [0.377]⋆⋆⋆	−0.835 [0.595]	0.208 [0.177]	0.429 [0.115]⋆	0.564 [0.254]⋆⋆⋆	0.537 [0.417]	0.429 [0.115]⋆	0.564 [0.254]⋆⋆⋆
Union density	−0.078 [0.039]⋆⋆⋆	−0.125 [0.038]⋆	−0.19 [0.039]⋆	−0.048 [0.052]	−0.131 [0.075]	−0.192 [0.099]⋆⋆⋆	−0.083 [0.037]⋆⋆	−0.121 [0.036]⋆	−0.124 [0.057]⋆⋆⋆	−0.23 [0.147]	−0.121 [0.036]⋆	−0.124 [0.057]⋆⋆⋆
Minimum wage/mean wage	−0.064 [0.043]	−0.105 [0.041]⋆⋆	−0.136 [0.041]⋆⋆	−0.074 [0.065]	−0.157 [0.065]⋆⋆	−0.193 [0.093]⋆⋆⋆	−0.045 [0.040]	−0.072 [0.034]⋆⋆⋆	−0.088 [0.052]	−0.104 [0.144]	−0.072 [0.034]⋆⋆⋆	−0.088 [0.052]
Passive labor market policy/gdp (×100)	−0.04 [0.010]⋆	−0.036 [0.010]⋆	−0.055 [0.034]	−0.044 [0.015]⋆	−0.04 [0.016]⋆⋆	−0.089 [0.067]	0.025 [0.009]⋆⋆	0.022 [0.010]⋆⋆	0.045 [0.034]	−0.024 [0.027]	0.022 [0.010]⋆⋆	0.045 [0.034]

Employment protection legislation [1–6]		−0.022 [0.010]★★	−0.023 [0.016]		−0.041 [0.022]★★	−0.047 [0.031]		−0.029 [0.006]★	−0.031 [0.007]★		−0.029 [0.006]★	−0.031 [0.007]★
Enrollment rate in early child care and preprimary education		−0.111 [0.032]★	−0.174 [0.042]★		−0.139 [0.064]★★	−0.186 [0.090]★★★		−0.019 [0.037]	−0.002 [0.057]		−0.019 [0.037]	−0.002 [0.057]
Agreements coverage			0.052 [0.056]			0.064 [0.100]			0.051 [0.044]			0.051 [0.044]
Bargaining centralization			−0.056 [0.042]			−0.052 [0.091]			−0.012 [0.050]			−0.012 [0.050]
Active labor market policy/gdp (×100)			0.015 [0.047]			0.078 [0.103]			−0.042 [0.051]			−0.042 [0.051]
Social expenditure/gdp			0.891 [0.937]			−0.632 [1.855]			0.22 [0.711]			0.22 [0.711]
Tax wedge			0.059 [0.091]			−0.031 [0.192]			−0.121 [0.117]			−0.121 [0.117]
Unemployment benefit Replacement rate			0.039 [0.074]			0.042 [0.105]			−0.022 [0.052]			−0.022 [0.052]
Observations (countries)	29	23	23	29	23	23	29	23	23	29	23	23
R-squared	0.66	0.78	0.84	0.46	0.69	0.74	0.47	0.71	0.81	0.31	0.71	0.81

Note: unless specified, all variables are in percent points—★significant at 1%; ★★significant at 5%; ★★★significant at 10%—robust standard errors in brackets—constant included.

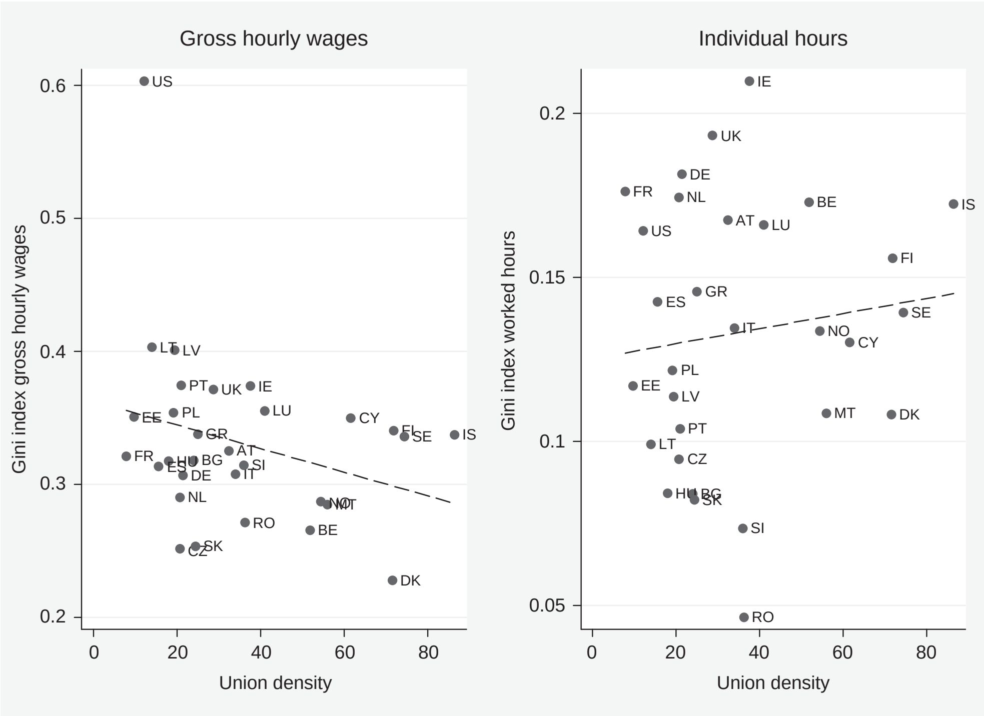


Figure 18.31 Earnings inequality (SILC 2010–PSID 2011) and union density (average 2001–2010).

labor force is negatively and significantly associated with inequality. Similarly, it occurs for wages, but not for hours: not surprisingly, when more women enter the labor market, the working hours regime as a whole becomes more diversified.[141]

When we introduce institutional measures to capture deviations from market equilibrium, we identify a subset of institutions that are significantly correlated with different inequalities (see columns 2–5–8 of Table 18.6). Union density has a negative association with yearly earnings, hourly wages, and hours: this captures different dimensions of union presence (such as coverage or wage centralization, which are not statistically significant[142]). Although the unconditional correlation with worked hours appears positive (see Figure 18.31), once we control for compositional effects it turns negative (despite a rather small magnitude). A second institutional dimension with a statistical negative correlation with earnings inequality is the presence and the level of minimum wages. However, as discussed in Appendix C, this institution is present only in a subset of

[141] Additional compositional controls related to the age composition do not come out statistically significant and therefore are left out of the analysis.

[142] Also strike activity is not statistically significant, but in addition it reduces the sample to 18 countries, and therefore is not shown.

countries, while in others this role is played by legislative or judicial extension of the union-bargained wage. In addition, there are often derogations for marginal workers, which are not captured by this measure. Nevertheless, the mere existence of a legal floor to downward flexibility of wages contributes to the containment of inequality.

The third institutional dimension deals with unemployment benefit, whose theoretical expectation is ambiguous due to a potential enhancing effect on the unemployment rate. The replacement rate does not exhibit a statistically significant correlation, whereas the overall public expenditure on passive labor market policies is negatively correlated with earnings and wage inequalities, and positively with hours inequality.[143] This suggests that transferring money to members of the labor force (which constitutes our sample of investigation) reduces inequality in terms of revenues, but on the other side allows for the continuation of unequally distributed job opportunities. A fourth dimension is connected to the employment protection.[144] Not surprisingly, its correlation is strongest with the distribution of work: the more regulated the labor contract, the more equal is the distribution of worked hours. Because employment protection and union activities tend to be complements (Bertola, 2004), it is not surprising to find an analogous negative correlation with earnings and wage inequality, as clearly shown in Figure 18.32.

Still restricting our examination to the subsample of OECD countries, we find some statistical evidence of a negative correlation of earnings inequality with child care attendance, interpreted as a proxy for child care availability. On a theoretical ground, we do expect a larger female participation in the labor market and an evener distribution of external work opportunities in the couple: both should have an impact on the hours inequality, which, however, do not appear in the data. The negative correlation with earnings inequality could capture some unobservable dimension of welfare provision, which is typically associated to lower inequality (though a direct measure of it, given by social expenditure, does not come out statistically significant).[145]

Despite the limited degrees of freedom, these are the only institutional features that correlate with statistical significance with various dimensions of earnings inequality. Against the potential objection of omitted variables, we have also introduced all measures that we have collected (see columns 3–6–9 of Table 18.6), without finding any other statistical correlation. However, despite the richness of the institutional framework, a simple cross-country regression such as the actual one does not provide an incontrovertible evidence of LMIs contributing to shape earnings inequality. To this end, we now move to exploit cohort variation in inequalities.

143 Data on the expenditure on labor market policies are not available in the case of Iceland.

144 The OECD measure of EPL is not available for non-OECD members (Bulgaria, Cyprus, Latvia, Lithuania, Malta, and Romania). However, in order not to lose these countries in the analysis of other institutions, we have imputed these missing values using the sample mean of nonmissing countries.

145 If we reduce the number of countries even further (to 21) by introducing measures of parental leave, we find some statistical significance for a negative correlation with inequality in hourly wages (not shown).

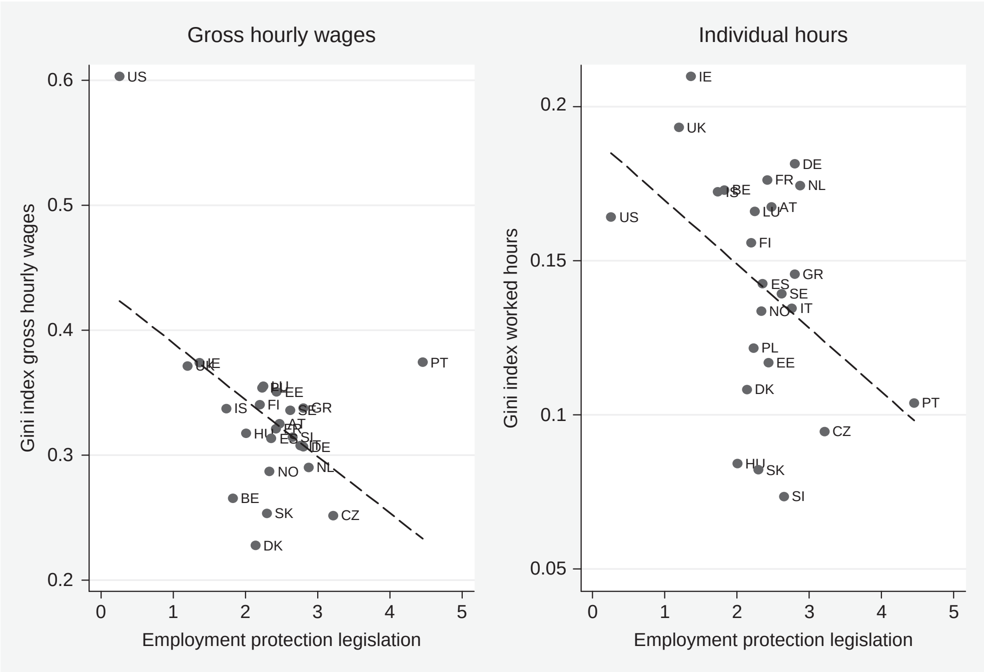


Figure 18.32 Earnings inequality (SILC 2010–PSID 2011) and employment protection (average 2001–2010).

In columns 10–11–12 of Table 18.6 we have considered as a dependent variable the correlation computed at the country level between hourly wages and worked hours, following the idea that higher correlation (in absolute terms) may reduce earnings inequality (as long as this correlation does not simply capture spurious correlation—see again Figure 18.22 and the discussion there). We find a negative correlation with both union density and employment protection legislation, suggesting that in a highly regulated labor market (due to firing restrictions and/or active union presence) the working poor obtain partial compensation of their weak command in the labor market by extended (or just complete) working hours.

18.5.3.2 Longitudinal or Pseudo-Longitudinal Approach

Aiming to obtain more statistically robust results, we need to exploit cross-country and within-country variations of inequality and institutions, to be able to dispense with unobservables by means of appropriate country and time-fixed effects. If data were available, one could take repeated cross sections for each country, compute inequality

Table 18.7 Matching rules between inequality measures and institutional variables

Cohort	Individual birth year	Age in 2010	Matching rule 1a: average institutional measures prevailing when entering the labor market aged 20-year old	Matching rule 1b: average institutional measures prevailing just before the entrance in the labor market (5-year lag)	Matching rule 2: average institutional measures prevailing over the entire working life course
1	1986–1990	20–24	2006–2010	2001–2005	2006–2010
2	1981–1985	25–29	2001–2005	1996–2000	2001–2010
3	1976–1980	30–34	1996–2000	1991–1995	1996–2010
4	1971–1975	35–39	1991–1995	1986–1990	1991–2010
5	1966–1970	40–44	1986–1990	1981–1985	1986–2010
6	1961–1965	45–49	1981–1985	1976–1980	1981–2010
7	1956–1960	50–54	1976–1980	1971–1975	1976–2010
8	1951–1955	55–59	1971–1975	1966–1970	1971–2010
9	1946–1950	60–64	1966–1970	1961–1965	1966–2010

measures of the relevant population in each survey, and match them with the prevailing institutional measures. Unfortunately, cross-country comparable surveys for the countries under analysis do not go back more than a couple of decades, and this has led us to pursue an alternative strategy. Because we need to match individuals belonging to different age cohorts, who entered the labor market in different years, to institutional profiles that are relevant for their wage determination, we need to discuss the appropriate matching rule.

One possibility would match individuals to the institutions prevailing at the time of their entrance into the labor market (see matching rules 1a and 1b in Table 18.7). This implies that the current difference between a person's wage and the wages of his or her coworkers may be affected by the bargaining activity exerted by the unions 30 years ago. As long as wages are highly persistent (due to seniority rules and/or automatic adjustment clauses) this may be considered a viable assumption. An alternative possibility considers both institutional persistence (institutions are slow-changing variables) and different exposure to an institutional environment (variable treatment). In this second perspective, older individuals are supposed to have been exposed to an institutional framework that has been (on average) available over their entire working life (see matching rule 2 in Table 18.7). In such a case, the current difference between someone's wage and the wages of his or her coworkers has been affected by the bargaining activity exerted over the past 30 years. To appreciate differences in the institutional measures according to the different matching rules, Figure 18.33 plots the

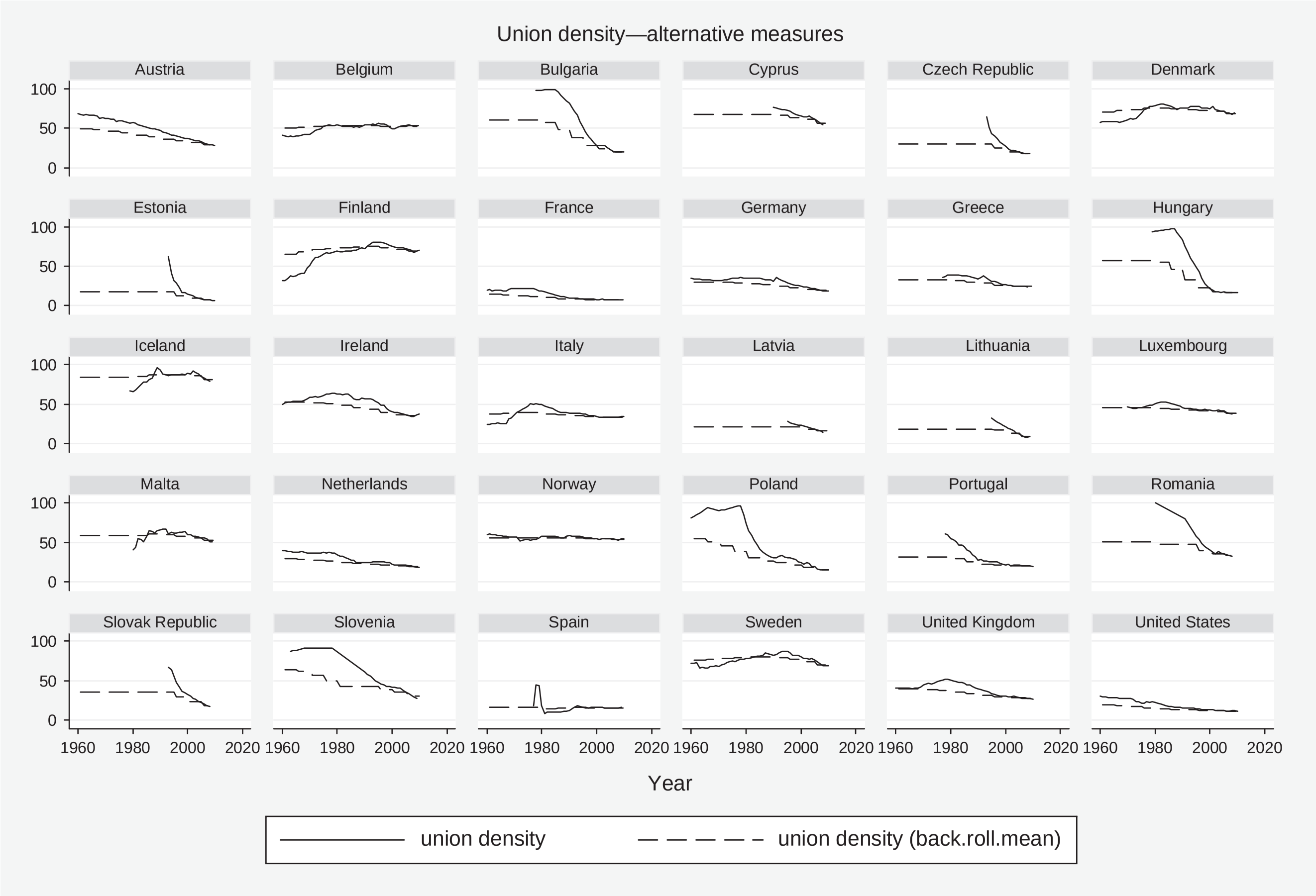


Figure 18.33 Alternative measures of exposition to institutions: union density.

contemporaneous union density (solid line) and the backward (moving) mean according to the third matching rule (dashed line): although the former is more volatile, the latter "keeps" a smoothed memory of past dynamics.

Both strategies are approximations because they induce measurement errors in the dependent variables (measuring wage inequality by age cohort is used as proxy for overall inequality measured in the past). However, they have the advantage of covering a long time span, allowing greater variability in the institutional measures.

Irrespective of the chosen matching, by treating our cross section as a pseudo-panel we significantly augment the degrees of freedom in the estimation. The different time coverage of institutional measures yields an unbalanced panel, where we control for country and cohort fixed effects. The errors are clustered at the country level. As a consequence, our results are more robust than the previous cross-section estimates reported in Table 18.8. As long as the fixed effects clean away all the other sources of confounding variations, we use cross-country and life-cycle variations in inequality for identifying the contribution of institutions to shape the earnings distribution. The contemporaneous insertion of several institutional measures allows for the identification of each specific contribution, other institutions and sample composition kept constant. We have decided to exclude the two oldest cohorts, inasmuch as information on institutions in the 1960s is available only for union density and unemployment benefit. In addition, retirement rules vary across countries, introducing large variations in the employment rate for these age cohorts.[146]

In Table 18.8 we present the estimates corresponding to the matching rule 1a (individual matched to the institutions prevailing when entering the labor market—the other matching rule 1b gives similar results on a shorter sample size, and is not reported for brevity). The structure of Table 18.8 resembles the previous Table 18.6 but leaves out the analysis of the correlation between hours and wages. We consider three measures of inequality (yearly earnings for full-time workers, hourly wages, and hours worked) and for each of them we control for educational attainment in the labor force and female participation. In both cases they exert a negative impact on inequality, despite the weaker statistical significance of education. For each dependent variable we consider three specifications: country fixed effects (columns 1–4–7), country and cohort fixed effects (columns 2–5–8), and country and cohort fixed effects including OECD indicator for employment protection, which excludes non-OECD members (columns 3–6–9).[147]

[146] The employment rate for individuals aged 55–64 ranges from 65% in Sweden (or 62% in the United States) to 30% for Italy or Romania.

[147] The first two columns still exclude Iceland, due to the lack of data on labor market policies, while the third excludes non-OECD countries.

Table 18.8 Gross earnings inequality (SILC 2010–PSID 2011) against market and labor market institutions (1975–2010) effects—OLS—longitudinal cohort data (matching rule 1a)

	1	2	3	4	5	6	7	8	9
Dependent variable	**Gini of gross yearly employee earnings (working more than 1000 h per year)**			**Gini of gross hourly wages**			**Gini of employee hours (all durations)**		
Population share with secondary degree	−0.022 [0.165]	0.041 [0.140]	0.028 [0.138]	−0.074 [0.154]	−0.075 [0.159]	−0.093 [0.160]	0.082 [0.104]	0.089 [0.109]	0.093 [0.120]
Population share with postsecondary degree	−0.162 [0.068]*	−0.104 [0.090]	−0.127 [0.091]	−0.18 [0.069]*	−0.239 [0.100]*	−0.285 [0.097]**	−0.084 [0.065]	−0.093 [0.108]	−0.103 [0.101]
Female employment rate	0.309 [0.141]*	−0.01 [0.169]	−0.035 [0.167]	0.315 [0.203]	0.333 [0.211]	0.318 [0.223]	−0.452 [0.121]**	−0.462 [0.202]*	−0.472 [0.220]*
Union density	0.138 [0.055]*	0.053 [0.045]	0.051 [0.046]	0.09 [0.041]*	0.073 [0.060]	0.058 [0.062]	−0.058 [0.033]***	−0.063 [0.037]	−0.076 [0.041]***
Minimum wage/ mean wage	−0.127 [0.073]***	−0.136 [0.055]*	−0.149 [0.063]*	0.064 [0.055]	0.066 [0.062]	0.059 [0.070]	−0.053 [0.028]***	−0.049 [0.030]	−0.04 [0.031]
Passive labor market policy/gdp	−0.004 [0.005]	−0.009 [0.006]	−0.007 [0.006]	−0.011 [0.007]	−0.012 [0.008]	−0.009 [0.008]	0.008 [0.004]*	0.007 [0.005]	0.008 [0.005]***
Employment protection legislation			0.026 [0.022]			0.011 [0.026]			−0.022 [0.018]
Observations	130	130	113	130	130	113	130	130	113
Number of countries	29	29	23	29	29	23	29	29	23
R-squared	0.4	0.5	0.56	0.19	0.22	0.28	0.64	0.65	0.67
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort fixed effects	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes

Robust standard errors clustered by countries in brackets—; *significant at 5%; **significant at 1%; ***significant at 10%.

In this framework we find only partial support to our previous findings with cross-sectional analysis. Focusing on a model that includes both country and cohort fixed effects, there is some evidence of a negative impact of unions on the distribution of work (column 9) and of a stronger impact of the minimum wage on earnings inequality. Contrary to previous results, passive labor market policies do not reach statistical significance for their negative impact on earnings and wage inequality, but register some positive impact on the Gini index for hours worked. Other institutional variables (such as the tax wedge, unemployment benefit, parental leave, and active labor market policies), which are constantly nonsignificant are not reported for brevity.[148] The same results are reinforced when we adopt the second matching rule, as shown in Table 18.9. The different data organization significantly extends the sample, and this allows for a more precise identification of the effects (see, for example, the unconditional correlation with passive labor market policies, depicted in Figure 18.34). Union density is now clearly reducing inequality in hours, and the minimum wage reduces inequality both in earnings and hours. In addition to the negative contribution of passive labor market policies on earnings and wage inequality, we now find that also active labor market policies negatively contribute to inequality reduction, possibly owing to the reduction in unemployment (i.e., more workers become employed earning a wage higher than the benefit).

18.5.4 Discussion

Our empirical results are consistent with the main findings in the literature reviewed in Section 18.4.6.[149] They confirm that the presence and stringency of a minimum wage reduces earnings inequality, also setting an (implicit) control on the distribution of working hours, which seems to be the main channel of inequality reduction of the bargaining activity of unions. Less common in the literature is the finding of a negative impact of both active and passive labor market policies. Here, we surmise that most of this effect works through variations in the unemployment rate: when active labor market policies are effective in pushing the unemployed back to work (at least for some hours) they reduce the bottom tail of the earnings distribution; when the unemployment support becomes more generous and/or more universal (as has happened during the current recession) it reduces the income gap between employed and unemployed, but potentially

[148] The other institutional measures appearing in Table 18.7 and not in Table 18.9 (child care, social expenditure, tax wedge) are not reported because they are not available over a longer time span going back to the older cohorts.

[149] Issues of data quality and a review of main findings for cross-country analysis can be found in Eichhorst et al. (2008).

Table 18.9 Gross earnings inequality (SILC 2010–PSID 2011) against market and labor market institutions (1975–2010) effects—OLS—longitudinal cohort data (matching rule 2)

	1	2	3	4	5	6	7	8	9
Dependent variable	**Gini of gross yearly employee earnings (working more than 1000 h per year)**			**Gini of gross hourly wages**			**Gini of employee hours (all durations)**		
Population share with secondary degree	0.018 [0.102]	0.1 [0.089]	0.053 [0.118]	−0.019 [0.089]	0.012 [0.099]	−0.017 [0.137]	0.097 [0.070]	0.092 [0.075]	0.104 [0.088]
Population share with postsecondary degree	−0.12 [0.051]*	−0.052 [0.065]	0 [0.061]	−0.116 [0.047]*	−0.104 [0.060]**	−0.091 [0.076]	−0.053 [0.043]	−0.013 [0.078]	−0.039 [0.095]
Female employment rate	0.283 [0.115]*	−0.116 [0.140]	−0.09 [0.149]	0.314 [0.129]*	0.177 [0.163]	0.218 [0.194]	−0.409 [0.090]***	−0.285 [0.148]**	−0.356 [0.158]*
Union density	0.094 [0.064]	−0.053 [0.035]	−0.042 [0.053]	0.109 [0.053]**	0.067 [0.056]	0.083 [0.088]	−0.137 [0.041]***	−0.082 [0.038]*	−0.123 [0.037]***
Minimum wage/mean wage	−0.226 [0.127]**	−0.22 [0.095]*	−0.259 [0.098]*	0.053 [0.101]	0.063 [0.098]	0.051 [0.105]	−0.102 [0.058]**	−0.119 [0.070]**	−0.101 [0.079]
Active labor market policy/gdp	−0.094 [0.032]***	−0.101 [0.024]***	−0.102 [0.024]***	−0.113 [0.035]***	−0.119 [0.037]***	−0.119 [0.036]***	−0.043 [0.028]	−0.049 [0.026]**	−0.051 [0.024]*
Passive labor market policy/GDP	−0.009 [0.020]	−0.037 [0.014]*	−0.047 [0.014]***	−0.054 [0.017]***	−0.064 [0.017]***	−0.068 [0.020]***	0.007 [0.013]	0.007 [0.013]	0.011 [0.014]
Employment protection legislation			0.035 [0.053]			−0.007 [0.062]			−0.057 [0.046]
Observations	203	203	161	203	203	161	203	203	161
Number of countries	29	29	23	29	29	23	29	29	23
R-squared	0.33	0.48	0.56	0.33	0.37	0.4	0.61	0.65	0.67
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort fixed effects	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes

Robust standard errors clustered by countries in brackets—*significant at 5%; **significant at 10%; ***significant at 1%.

Figure 18.34 Earnings inequality (SILC 2010–PSID 2011) and passive labor market policies, 5-year averages (1975–2010).

raises the unemployment rate. The combined effect of these channels seems to be overall inequality-reduction.

18.6. CONCLUSION AND FUTURE RESEARCH

Putting the literature on the distribution of (individual) wages in the context of the (household) income distribution we are struck by the disconnect between the two. There is an extensive literature with a long tradition on each of them but very little on both, despite the fact that wage earnings are by far the most important source of income in modern society. The strong shift from single-earner to dual-earner households that has come about with the rapid growth of female and in many cases part-time employment and the growing attention paid to the phenomenon of household joblessness make this an important lacuna. Significant policy implications may be suspected. The debate on household joblessness has already put into question the workings of labor market policies. The important debate on job polarization ignores how households may be affected or, alternatively, how they may offer compensation for the process. The interlinking of the two distributions raises doubts about the policies of redistribution. Traditional instruments found on each of the two sides, such as the minimum wage or income taxation, expectedly work out differently in a dual-earner world where household labor supply can involve low-paid jobs or low (part-time) earnings even at high levels of household income. As a result, the effects of these instruments will change and their political support in society may also be altered. Future research will require improved and systematic fact finding, the analysis will generate a better understanding of earnings as well as hours of work on an *annual* basis, consistent with incomes, and also broaden to include other, often newly minted institutions that affect joint household labor supply, such as child care provisions. There is no shortage of detailed research on various issues; however, the broad picture of the distributions as such is lacking. Connections run in both directions, from earnings to incomes as much as the other way around, and household formation and concomitant household labor supply cannot be taken as a given but are affected by both. Importantly, often the national work force has also doubled over recent decades, as a result of rapidly rising educational attainment and female labor market participation, and not only the global work force after the demise of communism.

Subsequently turning to the distribution of wages alone we have gone back to the origins of the debate in the early 1980s and sketched developments toward what is now a large and complex literature. We find that the unanticipated rise in earnings inequality in the United States over the 1970s put LMIs, such as (declining) unionization, as one possible explanation among others, such as demography or deindustrialization, on the research agenda. During the 1990s, the debate gave rise to the thesis of skill-biased technological change but also to international comparisons. The former approach has focused on market forces of supply and demand, the latter deemed those insufficient because of the growing international divergence in wage-inequality trends and has put the limelight on the role of national LMIs. After some leapfrogging of the two approaches from one consensus explanation to another during the 1990s, the two seem to be increasingly growing apart during the 2000s when important new contributions were made to the disadvantage of an integrated approach that could give each its proper place. Both sides may be at risk of creating "Ptolemaic epicycles" aimed to incorporating new observations. The supply-and-demand approach is challenged by the need for finding better empirical measures of technological change aimed at dispelling doubts that the relative demand of skills may be a tautology. It has added "tasks," "offshoring," and even "consumer preferences," which risk being ad hoc additions,[150] without realizing their

[150] Compare Autor's (2013, 25) remark "that there are almost as many distinct task classifications as there are papers in the task literature."

institutional preconditions. The institutional approach, on the other side, faces an abundance of institutions and ever new ones are added. It lacks a sufficiently clear-cut concept of institutions, ranging from laws, regulations, and habits to actual policies, and of their interactions—be they mutually reinforcing or compensating—on the one hand, and a clear criterion for delineating the institutional scope on the other hand.

In light of this, the double aim of future work on both sides should be to foster itself and to integrate the other side into its own framework at the same time. Pursuing this may seem more a problem of empirical method for the institutional side, although by contrast on the demand and supply side the problem may be more one of theoretical method as long as institutions keep being viewed as bodies alien to the market and to theorizing. For both sides there is a perspective of work to do at the firm level. Matched employer–employee data can help enlighten the role of both institutions as well as labor supply and demand. Such data are increasingly becoming available. This brings us back to the availability and quality of the data in addition to the earnings/incomes fact-finding already mentioned. Data and analysis shall move beyond the commonly used earnings data for full-time workers only, which are less and less representative especially at the margins of earnings and incomes. Therefore, more needs to be done regarding individuals' and households' work efforts and earning outcomes on an annual basis. In addition, though much attention has been paid to data quality, a better grasp of the customary use of inequality measures—currently, each of the two approaches has its own rather exclusive preferences—seems desirable.

Finally, we have set up a simple model accounting for the correlation of the different components of inequality (between and within) with LMIs. We find indeed that unionized labor markets are ceteris paribus less unequal in terms of annual earnings, because both hourly wages and worked hours are more evenly distributed. We improve on existing approaches with the help of a pseudo-longitudinal approach linking workers cohort-wise to the change in institutions over their working life in three different ways. Empirical results of three cross-country exercises focusing on different inequality measures and covering the United States and all European countries in 2010–2011, suggest inequality-reducing effects of unionization for hours, and of minimum wages for both hours and earnings.

ACKNOWLEDGMENTS

We are grateful to Christina Haas, John Schmitt, Thomas Lemieux, Nicola Fuchs-Schündeln for their help with the data and Anna Salomons for some useful suggestions. We thank in particular our editors, Tony Atkinson and François Bourguignon, for their very helpful comments and suggestions.

APPENDIX A. COUNTRY CODES

ISO alpha-3 and alpha-2 country codes

AUS	AU	Australia
AUT	AT	Austria
BEL	BE	Belgium
BGR	BG	Bulgaria
CAN	CA	Canada (BC: British Columbia, ON: Ontario)
CHE	CH	Switzerland
CYP	CY	Cyprus
CZE	CZ	Czech Republic
DEU	DE	Germany
DEU-W	DE-W	West Germany
DNK	DK	Denmark
ESP	ES	Spain
EST	EE	Estonia
EU		European Union
FIN	FI	Finland
FRA	FR	France
GRC	GR	Greece (also named EL by Eurostat)
HUN	HU	Hungary
IRL	IE	Ireland
ISR	IL	Israel
ITA	IT	Italy
JPN	JP	Japan
KOR	KR	Korea
LTU	LT	Lithuania
LUX	LU	Luxemburg
LVA	LV	Latvia
MLT	MT	Malta
NLD	NL	Netherlands
NOR	NO	Norway
NZL	NZ	New Zealand
POL	PL	Poland
PRT	PT	Portugal
ROM	RO	Romania
SWE	SE	Sweden
SVK	SK	Slovak Republic
SVN	SI	Slovenia
UK	UK	United Kingdom (official code GBR not used)
USA	US	United States

APPENDIX B. DATA SOURCES AND ADDITIONAL TABLES ON EARNINGS

We obtain data from the EUSILC survey conducted in 2010 (ver.1 dated 01/03/12), to which we added populations from Cyprus and Ireland extracted from the 2009 survey (ver.2 dated 01/08/11). Overall among European countries we consider 476,265 observations (of which 9902 Irish and 7557 Cypriots). For these individuals we know relevant demographics (age, gender, education,[151] marital status, birthplace), occupational characteristics (whether employed or self-employed, full-time or part-time, permanent or temporary contract, ISCO occupational code, workplace size, work experience).

We obtain data on the United States from participants to PSID survey conducted in 2011. Basic demographics (age, gender, education,[152] marital status, labor market status) are obtained from individual file (file ind2011er.zip downloaded on 22/07/13), which includes 24,661 observations. Information on labor earnings are collected from one respondent for each of 8907 households (typically a male household head), who responds about wage and hours for himself and his spouse (file fam2011er.zip downloaded on 22/07/13).

We adopt two selection rules:

(a) population in relevant working age, which we define as being between 20 and 64 years old. This is justified to allow for secondary school completion, and to take into account different early retirement rules in different countries. This leads to the exclusion of employed youngsters aged 15–19, a fraction of which is employed with an average hourly wage that is on average half of the average wage in the adult population. Because countries differ in the duration of compulsory education as well as in institutional design, we have preferred to leave the youngster component of the labor force out of our analysis.[153]

(b) population in the labor market, who self-define as either employed (employee or self-employed) or unemployed. This takes as exogenously given the significant cross-country differences in participation/employment rates (see Table 18.A1). These differences are even enhanced when we consider analogous rates computed at household level. Notice the high share of self-employed in Italy and Greece, which are also the countries where the share of top incomes accruing to them is

[151] Data on years of education have been computed from maximum educational attainment according to ISCED classification (variable PE040 in SILC) converted into years by using legal duration.

[152] Data on completed years of education are directly reported by the interviewees (variable ER34119 in PSID).

[153] There is an additional reason for excluding these cases as we identify cohabiting couples by taking the two first working members in the household. Retaining these individuals would increase the risk of mixing couples up with single earners and an earning child.

largest. The final sample is made of 264,216 individuals in the labor market, among which 201,500 employees, 33,384 self-employed, and 29,332 unemployed.

The labor earnings variable is defined as either "gross yearly earnings from dependent employment—cash or near cash"[154] or "gross yearly cash benefits or losses from self-employment" (see Table 18.A2 for means and standard deviations).[155] Unemployed subsidies received by (temporarily) unemployed workers are also considered in the computation of earnings inequality.[156]

In order to distinguish between annual earnings and hourly wages, we need information about the number of hours worked. In both survey hours worked are reconstructed thanks combining answers to two questions (weekly hours usually worked in recent months—thus referred to the period of interview—and number of months worked in the previous year).[157] The gross hourly wage rate is then computed dividing the yearly earnings by the hours worked.[158] Descriptive statistics on hours worked and hourly wages are reported in Table 18.A3. Notice that there is a significant loss of information when moving from yearly data (259,500 observations with nonnegative annual earnings) to hourly wage (228,153 observations with nonmissing hourly wages), due to missing information about weekly hours worked. A probit estimate indicates that young uneducated women holding a temporary contract are more likely not to report hours.

[154] Our GW variable (earnings from dependent employment) and GSELFW (earnings from self-employment) correspond to PY010G and PY050G variables, respectively, in EUSILC. In the case of PSID labor earnings are obtained from the sum of ER47501, ER47552, ER47582, and ER47612 variables (appropriately converted into yearly values) for the household head and from the sum of ER47752, ER47779, ER47809, and ER47839 variables for the working spouse. They are then separated between dependent employment or self-employment earnings according to the nonnegative value of the ER47495 or ER47752 variables ("how much is your salary").

[155] Negative values on earnings from self-employment are recoded into zeros, because most inequality indices (notably the Gini index) are defined over nonnegative values.

[156] This corresponds to the variable PY090G in EUSILC and to the variables ER48500/ER48619 (converted in annual values) in PSID.

[157] In EUSILC this corresponds to the variable PL060 (number of hours usually worked per week in main job) and the variables PL073-74-75-76-80 (number of months spent at full-time/part-time work as employee/self-employed (including family worker)/unemployed). In PSID this corresponds to the variables ER47456/ER47713 (On average, how many hours a week did (you/he/she) work on (all of) (your/his/her) (job/jobs) during 2010?) multiplied by variables ER47454/ER47711 (weeks employed last year—reconstructed variable from work histories) net of variables ER47633/ER47890 (weeks of vacation).

[158] In EUSILC, data on hourly wages are not fully temporally consistent, because the gross yearly wage and the months of work are referred to 2009, whereas the information about the weekly hours is referred to 2010. In PSID the interviewees directly provide a measure of hourly wage (variables ER47501/ER47758: What is your hourly wage rate for your regular work time?). In the case of the United States, where both measures are available, computed and elicited wages with positive values exhibit a correlation of 0.53.

Table 18.A1 Descriptive statistics computed from microdata—SILC 2010 and PSID 2011—labor market attachment (sample weights)

	Participation rate	Employment rate	Unemployment rate	Share self-employed	Female participation rate
Austria	0.734	0.676	0.079	0.124	0.596
Belgium	0.732	0.650	0.112	0.106	0.597
Bulgaria	0.800	0.670	0.163	0.096	0.620
Cyprus	0.754	0.710	0.058	0.139	0.644
Czech Republic	0.752	0.675	0.103	0.167	0.582
Denmark	0.782	0.715	0.085	0.094	0.702
Estonia	0.783	0.647	0.173	0.074	0.648
Finland	0.765	0.673	0.121	0.135	0.650
France	0.754	0.672	0.109	0.099	0.637
Germany	0.792	0.711	0.102	0.055	0.659
Greece	0.736	0.644	0.125	0.304	0.553
Hungary	0.687	0.595	0.134	0.124	0.542
Iceland	0.789	0.731	0.073	0.125	0.685
Ireland	0.707	0.592	0.162	0.149	0.532
Italy	0.682	0.607	0.109	0.219	0.485
Latvia	0.791	0.602	0.239	0.071	0.593
Lithuania	0.804	0.649	0.193	0.100	0.662
Luxembourg	0.739	0.692	0.063	0.078	0.609
Malta	0.632	0.593	0.063	0.136	0.423
Netherlands	0.760	0.732	0.037	0.154	0.660
Norway	0.797	0.771	0.033	0.073	0.741
Poland	0.716	0.640	0.105	0.214	0.571
Portugal	0.788	0.672	0.148	0.144	0.617
Romania	0.689	0.653	0.053	0.265	0.547
Slovak Republic	0.764	0.660	0.137	0.108	0.603
Slovenia	0.737	0.633	0.141	0.096	0.587
Spain	0.787	0.624	0.207	0.161	0.546
Sweden	0.823	0.776	0.057	0.043	0.751
United Kingdom	0.768	0.728	0.052	0.122	0.677
United States	0.791	0.709	0.104	0.109	0.683
Average	0.752	0.671	0.108	0.138	0.605

Table 18.A2 Descriptive statistics computed from microdata—SILC 2010 and PSID 2011–labor earnings (sample weights)

Country name	Gross earnings from dependent employment (mean)	Gross earnings from dependent employment (SD)	Gross earnings from self-employment (mean)	Gross earnings from self-employment (SD)	No. of observations with nonmissing values on yearly earnings
Austria	26,461.08	25,858.63	4096.55	17,355.84	6116
Belgium	26,723.36	20,772.17	2717.21	11,508.15	6387
Bulgaria	3024.31	2925.93	446.67	2378.81	7731
Cyprus	18,748.41	16,532.35	3139.14	13,762.04	3970
Czech Republic	7499.50	6937.02	1752.81	6380.98	9489
Denmark	38,504.27	26,550.36	2017.71	18,497.89	7005
Estonia	7507.27	6367.43	122.38	959.08	5891
Finland	27,382.17	21,351.46	2076.83	10,272.64	12,705
France	22,219.13	19,263.07	2110.23	15,184.50	11,518
Germany	24,586.15	22,526.16	1968.18	14,598.60	12,693
Greece	12,219.81	14,306.38	5645.76	18,567.37	7163
Hungary	5038.98	5062.23	777.84	3393.05	10,240
Iceland	23,832.86	17,620.38	818.24	3899.32	4075
Ireland	26,624.65	28,929.35	4520.33	19,799.64	4766
Italy	17,593.58	17,522.40	7017.56	24,495.35	19,637
Latvia	5942.19	6430.71	207.53	1370.13	6742
Lithuania	4979.52	5607.18	447.30	2589.25	6097
Luxembourg	42,588.71	38,248.88	2896.52	22,171.74	5717
Malta	13,787.59	11,406.58	2393.88	8493.19	3678
Netherlands	31,138.58	26,156.98	3186.18	16,507.76	11,621
Norway	42,686.63	32,290.34	3409.42	23,335.28	6269
Poland	5567.64	6017.03	1020.09	3613.49	14,693
Portugal	10,768.63	12,283.16	1432.51	6260.98	5655
Romania	2664.20	2452.14	333.46	1338.79	7342
Slovak Republic	6322.45	8057.32	683.42	2971.25	8071
Slovenia	14,493.82	12,900.63	1125.57	4699.40	14,085
Spain	14,620.40	14,235.10	1305.97	7504.83	16,812
Sweden	25,384.12	18,901.72	753.78	5208.48	8355
United Kingdom	24,787.20	28,023.56	3654.77	26,042.01	7818
United States	51,786.79	840,084.88	7103.13	35,807.86	7159
Average	18,819.88	32,903.78	2627.03	16,288.20	259,500

Note: Data in 2010 Euros except US where data are in 2011 US dollars.

Table 18.A3 Descriptive statistics computed from microdata—SILC 2010 and PSID 2011—hours and wages (sample weights)

	Hours worked (mean)	Hours worked (SD)	Hourly wage (mean)	Hourly wage (SD)	No. of observations with nonmissing values of hourly wages
Austria	1801.05	648.58	17.14	25.54	5689
Belgium	1742.40	630.42	18.74	16.46	5662
Bulgaria	1924.69	413.00	1.83	1.67	6381
Cyprus	1844.24	527.32	10.67	8.73	3743
Czech Republic	1979.71	489.53	4.23	3.29	8685
Denmark	1752.24	449.20	24.02	15.45	6615
Estonia	1761.99	498.25	5.22	5.93	4795
Finland	1718.94	558.07	19.46	21.05	11,479
France	1746.34	617.55	15.20	18.75	10,223
Germany	1746.93	619.15	15.24	11.98	11,569
Greece	1829.68	553.70	7.96	8.64	6212
Hungary	1844.58	403.42	3.05	2.71	8876
Iceland	1914.37	642.18	14.38	19.49	3783
Ireland	1637.80	693.63	22.17	29.44	3917
Italy	1829.60	474.38	11.17	10.77	17,248
Latvia	1816.58	487.95	4.07	4.06	5092
Lithuania	1784.31	415.07	3.38	3.41	4995
Luxembourg	1833.12	576.49	25.13	20.55	5311
Malta	1863.97	517.44	8.22	7.08	3353
Netherlands	1582.64	553.61	21.33	19.91	11,212
Norway	1765.72	504.90	25.81	21.14	6043
Poland	1894.65	517.62	3.40	4.36	13,077
Portugal	1866.69	470.88	7.02	7.65	4281
Romania	1950.27	352.44	1.46	1.24	6589
Slovak Republic	1913.54	395.13	3.81	4.11	7180
Slovenia	1897.84	415.46	8.76	7.16	12,131
Spain	1821.19	536.07	10.32	10.42	12,783
Sweden	1516.64	496.22	19.80	24.18	7755
United Kingdom	1785.60	637.22	15.39	19.19	7328
Unites States	1937.14	848.38	8.65	26.66	6146
Average	1793.65	564.99	12.37	15.44	228,153

Note: Data in 2010 Euros except US where data are in 2011 US dollars.

Table 18.A4 Estimates of model relevant parameters—employed or unemployed individuals—SILC 2010 and PSID 2011

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	Skilled workers	Unskilled workers	Unemployed	Average wage	Unemployment benefit	Replacement rate	Skilled wage	Unskilled wage	Wage premium	Between-group inequality	Overall inequality		
	$\alpha(1-u)$ pct. points	$(1-\alpha)$ $(1-u)$ pct. points	u pct. points	$\overline{w}$ euros/ dollars	b euros/ dollars	γ pct. points	w^s euros/ dollars	w^u euros/ dollars	σ pct. points	$Gini_{est}$	$Gini_{obs}$	$\Delta Gini$	$\frac{Gini_{est}}{Gini_{obs}}$
AUT	0.321	0.616	0.063	32,205.34	3948.59	0.12	41,021.68	27,604.24	0.49	0.073	0.348	0.28	0.21
BEL	0.439	0.461	0.100	33,006.56	5958.73	0.18	39,378.79	26,940.58	0.46	0.086	0.299	0.21	0.29
BGR	0.225	0.654	0.121	3789.24	449.98	0.12	5186.9	3309.31	0.57	0.100	0.346	0.25	0.29
CYP	0.391	0.567	0.042	21,922.22	2928.37	0.13	28,080.84	17,672.54	0.59	0.074	0.364	0.29	0.20
CZE	0.176	0.764	0.060	9785.79	926.04	0.09	14,118.73	8786.53	0.61	0.067	0.3	0.23	0.22
DEU	0.454	0.470	0.076	47,286.8	6110.35	0.13	55,418.81	41,882.96	0.32	0.067	0.387	0.32	0.17
DNK	0.377	0.566	0.057	8493.88	1626.26	0.19	10,215.24	7451.71	0.37	0.060	0.24	0.18	0.25
EST	0.322	0.532	0.145	34,994.66	4483.46	0.13	42,707.44	27,756.81	0.54	0.116	0.361	0.24	0.32
ESP	0.345	0.483	0.171	25,870.6	5178.83	0.20	32,662.33	21,905.82	0.49	0.119	0.352	0.23	0.34
FIN	0.434	0.463	0.103	30,659.61	3949.61	0.13	38,298.93	23,276.59	0.65	0.104	0.317	0.21	0.33
FRA	0.336	0.576	0.087	19,628.96	3255.67	0.17	23,998.9	16,440.65	0.46	0.080	0.321	0.24	0.25
GRC	0.381	0.522	0.097	6403.25	955.58	0.15	9143.06	5173.97	0.77	0.108	0.344	0.24	0.31
HUN	0.273	0.608	0.119	27,111.2	3831.86	0.14	32,224.25	23,284.92	0.38	0.089	0.366	0.28	0.24
IRE	0.484	0.361	0.155	38,583.85	8978.04	0.23	45,739.25	28,977.96	0.58	0.111	0.404	0.29	0.27
ISL	0.394	0.527	0.079	23,958.99	3936.05	0.16	29,992.45	22,250.91	0.35	0.069	0.326	0.26	0.21
ITA	0.203	0.718	0.079	7518.82	1252.23	0.17	10,476.27	5644.96	0.86	0.091	0.32	0.23	0.29
LTU	0.574	0.296	0.131	6335.65	961.26	0.15	7365.67	4337.72	0.70	0.108	0.424	0.32	0.26
LUX	0.296	0.649	0.055	44,751.21	11,201.46	0.25	67,207.76	34,529.19	0.95	0.095	0.372	0.28	0.25
LVA	0.314	0.495	0.192	16,856.53	3797.42	0.23	22,378.49	14,985.31	0.49	0.125	0.44	0.31	0.28
MLT	0.242	0.713	0.045	37,965.09	7090.32	0.19	47,164.16	30,699.66	0.54	0.061	0.305	0.24	0.20
NLD	0.432	0.547	0.021	48,789.02	6094.77	0.12	56,535.29	41,581.87	0.36	0.047	0.317	0.27	0.15
NOR	0.470	0.505	0.024	7390.9	996.99	0.13	9625.72	6253.77	0.54	0.062	0.316	0.25	0.20
POL	0.312	0.614	0.074	14,248.9	3843.58	0.27	26,901.73	11,502.8	1.34	0.122	0.362	0.24	0.34
PRT	0.159	0.734	0.106	3684.08	966.96	0.26	5262.04	3126.16	0.68	0.081	0.385	0.30	0.21
ROM	0.254	0.720	0.026	7954.01	1245.76	0.16	10,154.17	7159.13	0.42	0.046	0.27	0.22	0.17
SWE	0.426	0.530	0.045	18,068.09	2233.65	0.12	27,336.91	14,716.41	0.86	0.093	0.295	0.20	0.32
SVNI	0.242	0.669	0.089	20,305.54	4503.04	0.22	26,071.97	16,187.4	0.61	0.083	0.329	0.25	0.25
SVK	0.248	0.687	0.065	28,105.98	5413.02	0.19	31,987.71	24,987.56	0.28	0.051	0.29	0.24	0.18
UK	0.427	0.541	0.033	30,438.71	4831.86	0.16	39,293.29	23,445.94	0.68	0.076	0.404	0.33	0.19
United States	0.531	0.378	0.091	39,150.52	9903.6	0.25	47,253.16	27,769.21	0.70	0.090	0.408	0.32	0.22
Mean	0.342	0.571	0.087	23,162.53	4147.73	0.18	31,462.71	18,192.87	0.73	0.099	0.342	0.24	0.29

APPENDIX C. DATA SOURCES AND DESCRIPTIVE STATISTICS ON LMIs

Data on institutional measures were collected over a time interval spanning half a century, from 1960 to 2010.

Union density

It measures the fraction of wage and salary earners who are members of trade unions. It excludes unemployed and retired workers (net version). Source: ICTWSS database version 2 (Database on Institutional Characteristics of Trade Unions, Wage Setting, State Intervention and Social Pacts in 34 countries between 1960 and 2007—see Visser, 2009—variable UD—downloaded on 04/04/13).[159]

Coverage

It measures the fraction of employees covered by wage-bargaining agreements over all wage and salary earners in employment with the right to bargaining. Source: ICTWSS database version 2 (variable ADJCOV).

Wage centralization

It represents a summary measure (ranging between 0 and 1) of centralization and coordination of union wage bargaining, taking into account both union authority and union concentration at multiple levels—source: ICTWSS database version 2 (variable CENT).

Strike activity

It measures the days not worked for strikes and lockouts divided by participant worker—total economy. Source is ILO (downloaded on 04/04/13).

Minimum wage

It takes the ratio of the statutory minimum wage relative to mean wage of full-time workers (sometimes known as "Kaitz index"—see Dolado et al., 1996). However this measure does not consider the possibility of differentiation across workers types. For this reason, Aghion et al. (2011) have combined the ratio of the minimum wage to the GDP per capita with an index of stringency derived from ILO.[160] For this reason, the

[159] This measures highly correlates with the OECD corresponding measure (0.95) and with the ILO one (0.99), which is not surprising given the background studies conducted by the same author (Jelle Visser).

[160] This index takes value of 1 if there is a legal statutory minimum wage and if the minimum wage is set at the national level without any derogation, value of 0.5 if there is a legal statutory minimum wage but with derogations by age, qualification, region, sector, or occupation; or if the wage floor is set by collective bargaining but extended to all workers, and a value of 0 if the wage is set by collective bargaining and only applies to the unionized workers. This solution introduces a value of the index even if a country does not have the provision of a minimum wage, because otherwise these countries should be left out of picture. See the ILO TRAVAIL legal databases (http://www.ilo.org/dyn/travail), which however provides only the contemporaneous information (thus preventing us to use the measure for past periods).

variable is set to zero when minimum wage provision is absent. Data are downloaded from the OECD Stats website (except than in the case of Iceland, whose values are taken from Table 5.5 of Danish Technological Institute. *Assessment of the Labour Market in Iceland*. Contract no. VC/2010/038 Final report—Policy and Business Analysis—April 2011).

Employment protection legislation

The measure we use is provided by OECD, which recently have partially revised their country assessment (OECD, 2012).[161] It measures the stringency of firing regulation and is based on eighteen dimensions of the firing procedure.[162] There is a second series provided by World Bank, which has been used among others such as Botero et al. (2004).[163]

Unemployment benefit

Unemployment insurance and unemployment assistance benefits—gross replacement rate (ratio to the average wage) for a full-time adult worker. The source is OECD historical series, which is available in odd years and imputed using intermediate means in even years. It is the average between single worker and one-earner married couple with two children.[164]

[161] Data used in the analysis of the main text have been downloaded on 02/08/13. A preliminary download conducted on 4/4/2013 yields a different series for overall EPL, which however exhibit a correlation with the new one of 0.97.

[162] Eight dimensions concern "regular contracts": notification procedures, delay involved before notice can start, length of the notice period at various tenure durations, severance pay at various tenure duration, definition of justified or unfair dismissal, length of trial period, compensation following unfair dismissal, possibility of reinstatement following unfair dismissal. Six dimensions concern "temporary employment": valid cases for use of fixed-term contracts (FTC), maximum number of successive FTC, maximum cumulated duration of successive FTC, types of work for which temporary work agency (TWA) employment is legal, restrictions on number of renewals, maximum cumulated duration of TWA contracts. Four dimensions concern "collective dismissal": definition of collective dismissal, additional notification requirements, additional delays involved before notice can start, other special costs to employers—methodology is accurately described in chapter 2 of OECD, 2004)

[163] The World Bank index measures firing costs in terms of weeks of salary and it is based on three components: the notice period for redundancy dismissal, the severance pay for redundancy dismissal, and the legally mandated penalty for redundancy dismissal.

[164] It combines GRR(APW) until 2001 and GRR(AW) afterward. Eurostat provides a measure of the unemployment benefit net replacement rate for a single worker, which has a correlation index with OECD gross ratio equal to 0.55 and a limited time coverage, because it starts with 2001 and does not cover the United States.

Tax wedge

Average tax wedge (sum of social contributions and income taxes as ratio to the average wage). It considers the average between single worker with no child and one-earner married couple with two children. The source are the estimates from the OECD microsimulation model.

Social expenditure

It measures the expenditure for cash benefits and benefit in kind for social assistance, as percentage of GDP. The source is OECD historical series, which are available on 5-year base, and then interpolated.

Child care

It measures the enrollment rate in early child care and preprimary education (average between age 3, 4, and 5—full- and part-time students) and proxies the availability of child care facilities. Available values for years 2005 and 2010, while intermediate values are interpolated. The source is OECD, *Education at a Glance* 2012, Table C2.1.

Parental leave

It captures the possibility of reconciling work and fertility, by measuring weeks of paid leave for childbirth. The series is available since 1970. The source is Thévenon and Solaz (2013). Further documentation can be found at http://www.oecd.org/social/soc/oecdfamilydatabase.htm

Tax treatment of household incomes

This variable aims to capture the potential favorable tax treatment of working couples vis-à-vis individual taxation. It is constructed as the ratio between the average tax rate of single earner family (earning 170% of average wage) and the average tax rate of a two earners family (main earner at average wage and second earner making 67% of average wage). The reported variable consists of a further averaging between two household situation with respect to children (zero children and two children families). A higher value would indicate a favorable treatment of labor market participation of a second earner. Data available since 2001. The underlying data is obtained from the OECD microsimulation model, available at http://stats.oecd.org/Index.aspx?DataSetCode=FIXINCLSA.

Active and passive labor market policies

It considers the public expenditure on active or passive labor market policies as percentage of GDP. It combines two data sources: when available we have been using

OECD statistics for homogeneity with other series; otherwise we have resorted to Eurostat, which classifies as actives expenditure categories from 2 to 7 (2. Training—3. Job rotation and job sharing—4. Employment incentives—5. Supported employment and rehabilitation—6. Direct job creation—7. Start-up incentives) and passive expenditure categories 8 and 9 (8. Out-of-work income maintenance and support—9. Early retirement).

Overall means and standard deviations for these variables are reported in Table 18.A5. Country means are reported in table with reference to the most recent decade.

Table 18.A5 Descriptive statistics for institutional measures, sample period 1960–2010, 30 countries

Variable	No. of observations	Mean	SD	Minimum	Maximum
Union density	1123	45.99	22.97	6.67	100.00
Coverage	931	70.70	21.74	7.50	100.00
Centralization	1016	0.42	0.19	0.08	0.98
Strike activity	730	5.09	7.29	0.00	61.14
Minimum wage	1179	0.19	0.21	0.00	0.71
Employment protection legislation	506	2.37	0.88	0.26	5.00
Unemployment benefit	918	27.02	15.48	0.00	70.00
Tax wedge	294	25.32	7.75	8.17	41.88
Social expenditure	615	2.07	1.00	0.20	4.40
Child care	144	81.73	16.46	24.70	101.13
Parental leave	887	40.82	43.57	0.00	214.00
Tax treatment of household incomes	228	1.93	1.47	0.26	8.21
Active labor market policies	560	0.70	0.52	0.03	3.04
Passive labor market policies	572	1.25	0.99	0.08	5.45

Table 18.A6 Sample means of institutional measures—recent years (average 2000–2010)

	Union density	Coverage	Central-ization	Strike	Minimum wage	Employment protection legislation	Unemploy-ment benefit	Tax wedge	Social expenditure	Child care	Parental leave	Tax treatment of household incomes	Active labor market policies	Passive labor market policies
Austria	32.37	98.91	0.90	1.03	0.00	2.47	31.67	31.90	2.72	78.00	117.09	0.85	0.66	1.32
Belgium	51.88	96.00	0.46	23.07	0.44	1.82	40.91	36.97	2.63	99.54	28	1.93	1.17	2.23
Bulgaria	24.01	32.50	0.31	na	0.00	na	46.40	21.43	na	na	na	na	0.30	0.25
Cyprus	61.63	57.35	0.25	2.33	0.00	na	60.40	8.50	na	na	na	0.56	0.13	0.58
Czech Republic	20.74	43.77	0.25	na	0.31	3.21	6.14	17.74	1.87	82.01	185.64	0.33	0.23	0.27
Denmark	71.62	81.87	0.47	3.49	0.00	2.13	52.72	39.19	3.55	91.91	48.91	2.26	1.72	2.06
Estonia	9.69	25.06	0.36	0.21	0.30	2.43	50.00	17.60	1.79	87.08	na	0.67	0.11	0.39
Finland	71.84	87.13	0.40	2.16	0.00	2.19	34.56	32.57	2.98	51.29	42.35	2.69	0.90	1.84
France	7.84	90.00	0.21	22.47	0.47	2.42	39.88	24.90	3.02	100.99	32.55	0.69	1.02	1.46
Germany	21.44	64.40	0.48	1.60	0.00	2.80	25.85	36.73	1.96	91.57	60.1	4.11	1.03	1.78
Greece	25.03	65.00	0.34	na	0.33	2.80	14.42	22.77	1.14	72.30	24.09	1.06	0.17	0.44
Hungary	18.01	37.85	0.23	1.36	0.36	2.00	13.31	34.13	3.26	86.86	110	1.11	0.42	0.44
Iceland	86.47	88.02	na	16.74	0.51	1.73	40.76	26.00	3.00	95.44	25.97	3.16	na	na
Ireland	37.55	49.91	0.52	6.20	0.45	1.36	34.95	11.86	2.70	53.51	20.91	2.42	0.75	1.24
Italy	34.01	80.00	0.34	1.01	0.00	2.76	34.64	24.59	1.31	97.71	48	4.05	0.50	0.83
Latvia	19.41	19.41	0.48	4.01	0.33	na	60.00	25.11	na	na	na	1.08	0.19	0.46
Lithuania	13.93	12.16	0.30	2.41	0.36	na	47.62	23.61	na	na	na	2.74	0.18	0.23
Luxembourg	41.00	58.22	0.31	na	0.34	2.25	26.67	19.89	3.36	85.95	42	1.19	0.42	0.60
Malta	56.02	58.26	0.37	1.35	0.00	na	30.63	14.94	na	na	na	1.13	0.04	0.36
Netherlands	20.69	82.63	0.57	2.58	0.43	2.87	41.74	32.36	1.67	68.02	20.73	1.40	1.31	1.74
Norway	54.44	72.94	0.51	11.38	0.00	2.33	51.80	28.46	2.93	91.93	37.91	5.97	0.65	0.44
Poland	19.18	39.00	0.23	3.48	0.34	2.23	10.89	28.74	1.13	49.00	122.55	0.70	0.43	0.74
Portugal	20.92	60.07	0.34	1.35	0.36	4.45	42.23	18.55	1.15	81.19	18.65	1.75	0.64	1.11
Romania	36.28	70.00	0.25	13.27	0.29	na	32.20	27.34	na	na	na	na	0.08	0.38
Slovak Republic	24.34	44.70	0.50	0.01	0.35	2.30	9.50	18.22	1.89	72.53	164	0.99	0.30	0.46
Slovenia	35.96	97.40	0.40	na	0.43	2.65	61.64	29.09	1.07	80.67	na	1.00	0.29	0.41
Spain	15.60	87.84	0.36	2.55	0.34	2.36	34.98	16.41	1.16	98.68	16	0.65	0.79	1.78
Sweden	74.40	93.17	0.51	3.58	0.00	2.62	37.62	31.25	3.29	90.39	59.95	3.50	1.20	1.06
United Kingdom	28.74	34.64	0.11	2.20	0.36	1.20	15.99	25.87	3.14	91.88	28.55	1.84	0.36	0.23
United States	12.07	13.88	0.18	na	0.26	0.26	18.05	18.16	0.70	63.01	0	1.49	0.15	0.49

APPENDIX D. LITERATURE SUMMARY TABLES: HOUSEHOLD INCOMES AND EARNINGS AND WAGE DISPERSION AND INSTITUTIONS[165]

Table 18.A7 Household incomes and Earnings (see Section 18.2)

Authors	Years and countries	Data sets* and sample selection	Method and important variables	Main findings
i. Comparison of aggregate inequality measures				
Atkinson and Brandolini 2006	CAN, DEU, FIN, NLD, NOR, SWE, UK, United States; around 2000.	LIS; sample consecutively extending from employees aged 15–65 to all individuals in households.	International comparisons of earnings dispersion: cross section and trends. Comparison Gini's of wages, earnings of employed 15–64, earnings of all 15–64, income all 15–64, income all, equivalized income all.	They aim to take stock of existing research and conclude to, interalia: – need to model supply and demand and institutional variables in a common framework, linked to an underlying economic model; – differences in definitions and coverage may affect cross-country comparisons and cannot be assumed to be fixed over time. – need for care with different income concepts and different populations. – important interdependences between different explanatory variables.
Brown (1999)		Mainly US studies, from Gramlich 1976 to Neumark and Wascher 1997.	Several simple statistics: poor fraction among low-wage workers: around 20%. Card and Krueger (1995a) argue that other forces have increased the fraction. Probability of low pay among workers in low-income families. But many poor families have no workers. Gains of minimum wage increase comparable across deciles. Small impacts, difficult to find.	Effects of the minimum wage on the wage distribution became clearer with the declining real minimum wage in the 1980s; nevertheless the ability of minimum wages to equalize the distribution of family incomes remains quite limited.
Burtless (1999)	United States; 1979 vs. 1996.	CPS; ages 25–59.	Counterfactuals for Gini coefficient of household-equivalent personal income: holding constant male, female earnings distributions, or partners' earnings correlation.	Much of the rise in overall US inequality is due to family composition shifts and other causes rather than the change in pay patterns. Household income inequality change attributed to 33–44% earnings, shift to single and single-parent households 21–25%, partners' increased earnings correlation 13%.

[165] Tables A7 and A8 make generous use of summaries, abstracts, introduction and conclusions of the underlying papers.

Gottschalk and Smeeding (1997)	AUS, ISR, JPN, NOR, NZL, United States, and most of EU15; 1980s into early 1990s.	National data sets.	Comparison of studies of income inequality with discussion of roles of earnings, demography and social protection. The inclusion of multiple income sources received by multiple individuals thwarts attempts to identify the causal links that led to variations across time and across countries in the distribution of total posttax and transfer family income. Researchers have, therefore, limited themselves largely to purely accounting exercises which decompose changes in overall inequality into a set of component parts that may reflect endogenous as well as exogenous changes.	Better structural models of income distribution and redistribution that can be applied across nations are badly needed. Ideally, an overall framework would simultaneously model the generation of all sources of income (labor income, capital income, private transfers, public transfers, and all forms of taxation) as well as the formation of income sharing units. Although most of the components of such a model were identified as early as the mid-1960s, our progress toward building such a model has been slow. If we are to understand why we observe the extent and pattern of inequality levels and trends that are extant in this review, an overall conceptual framework with empirically testable components is the next big step that must be taken.
Gottschalk and Danziger (2005)	United States; 1975–2002.	CPS; ages 22–62 with positive earnings, males/females separately.	Comparison of wage rates: between/within inequality, and annual hours, family earnings, family incomes and equivalized incomes. No attempt to decompose the change in family income into its component parts because there are many ways to do so and there is no consensus on the most appropriate decomposition.	The similarity in the timing of changes in male wage rate inequality and family income inequality has been used as evidence that increased family income inequality primarily reflects increased inequality of wage rates. Authors show that other important factors were also at work. Female wage inequality actually declined steadily from 1975 through 2002. Although earnings inequality of males grew even more rapidly than wage inequality during the early 1980s, this largely reflects cyclical changes in hours. For females, changes in hours more than offset the rise in wage inequality. The acceleration in male wage and earnings inequality during the early

Continued

Table 18.A7 Household incomes and Earnings (see Section 18.2)—cont'd

Authors	Years and countries	Data sets* and sample selection	Method and important variables	Main findings
				1980s disappears when earnings of other family members are included. Thus, changes in work hours by other family members seems to have largely offset increased male labor market inequality.
Kenworthy (2008)	12: 9 old EU, United States, CAN, AUS; 1980–2005.	LIS for households, OECD for employment.	Cross-country comparison circa 2000 of pretax pretransfer equivalized household income inequality (P75:P25 ratios) to individual earnings inequality, (part-time) employment rate, zero-earner household rate, singles and marital homogamy, and to posttax posttransfer inequality (Gini's).	Thus, although much of the cross-country variation in levels of posttax–posttransfer income inequality is a product of differences in levels of market inequality, redistribution is also important. For understanding developments over time, redistribution is essential. Thus the focus ought to be chiefly on employment and redistribution, rather than on wage inequality and/or household composition.
Reed and Cancian (2001)	United States; 1969–1999.	March CPS; families with adults aged 25–59.	A new approach to measuring source contributions that has three advantages over inequality decompositions. First, a clear counterfactual, "What would have been the change in family income inequality were it not for the change in the distribution of the income source?" Second, simulation of counterfactual distribution of family income, allowing use of multiple summary measures of inequality and evaluation of impact at various points in the distribution (e.g., the 10th and 90th percentiles). Third, incorporate married-couple and single-person families and account for changes in marriage rate.	Changes in distribution of male earnings account for more of the growth in family income inequality than do changes in any other source of income. Changes in the distribution of female earnings have reduced family income inequality.

Salverda and Haas (2014)	EU (ex. CYP, MLT); 2010.	SILC 2011; employees in households with main income from earnings and working-age nonstudent head.	Decile comparisons with fixed household rankings of individuals; annual earnings with breakdown by hourly rates and annual hours worked; number of earners in household over deciles of household earnings.	Cross-country differences in individual labor-market earnings inequality are amplified by household combination of labor supply and the correlation of pay between its members. The former effect is always much larger (+99%) than the latter (+11%).
Večerník (2010)	CZE, HUN, POL, SVK, and AUT, DEU; from late 1980s on for CZE, HUN, POL, SVK, 2007 for all six.	LIS, and SILC for 2007; employees only.	Quintile shares and Gini's. Pearson coefficient correlations of household incomes with personal earnings for males/females; decomposition of Gini's.	Even with the best possible data on personal and household incomes available for analysis, there is still much we do not know about income sources, development and inequality. In fact, we cannot expect that income statistics will ever be capable of describing real incomes and income inequality in full. However, not having any other source of general information about income distribution, we cannot do anything else but examine the surveys from various angles and try, from time to time, to look beyond just data.
Večerník (2013)	CZE; 1988, 1992, 1996, 2002, 2009	National microcensuses and Czech part of SILC.	OLS regressions of contributions by sex, age and education to couples' earnings.	Increasing influence of education is the personal earnings of employees; in couples, education has an important impact on both women's employment and their earnings; the importance of marital partners' education levels on household income grew even more than its effect on earnings.
ii. Decompositions of incomes				
Brewer et al. (2009)	UK; 1968–2006.	HBAI; all individuals.	Decomposition: Shorrocks (1982) (Paul, 2004). Incomes: regression-based methodology developed by Fields (2003) and Yun (2006). Incomes: after all direct taxes and all state	Changes in within-group inequality are always the dominant explanatory factor in changes in overall inequality, although between-group effects also contribute significantly in some periods. Changes in relative incomes between groups are the

Continued

Table 18.A7 Household incomes and Earnings (see Section 18.2)—cont'd

Authors	Years and countries	Data sets* and sample selection	Method and important variables	Main findings
			benefits and tax credits. Individual earnings: gross.	major source of this between-group variation, though population changes also have a particularly significant impact in the early 1980s—presumably due to the rising number of workless households. The relative incomes of multi-earner households climbed steadily throughout almost the entire period we study. Income inequality: large unexplained residual term, even more so for change; employment status and occupation is by far the most significant explanatory variable, explaining almost a third of total income inequality in 1972. The residual is also important for earnings though less.
Brewer and Wren-Lewis (2012)	UK; 1968–2009.	HBAI; all individuals.	Three complementary decomposition methods: (1) Decomposition by income source, following Shorrocks (1982); (2) decomposition by population subgroup, following Mookherjee and Shorrocks (1982) and Jenkins (1995); (3) decomposition by factor, following Fields (2003). Because inequality in earnings (among individuals in employment) is an important source of changes in overall income inequality, the second and third decompositions are performed on individual earnings inequality, as well as on household income inequality.	Inequality in gross employment and self-employment income grew but since 1991 effect on inequality in total income almost entirely offset by: (1) declining inequality between those with different employment statuses, primarily due to a fall in unemployed people, (2) mitigation by employment taxes, (3) investment income became less unequal largely due to the decline in its importance, (4) rise in relative incomes of pensioners and households with children under 5.

Cancian and Schoeni (1998)	United States; 1968–1995.	March CPS; all persons related and residing are part of the same family, only families with prime-age heads (22–55). Exclude military, farmers, self-employed, students, and those living in group quarters.	Decomposing coefficient of variation. Estimate impact of wives' earnings using four alternative counterfactual reference distributions. If observed distribution of income is more equal than counterfactual distribution, then wives' earnings can be said to be equalizing.	Changes in husbands' earnings are substantially more important in explaining recent trends.
Cancian and Schoeni (1998)	AUS, CAN, FRA, DEU-W, NOR, ISR, SWE, CHE, UK, United States; 1980s.	LIS; husbands and wives.	Splitting CV2 when there are only two components of income: earnings of husband and of wife, into parts. Interest in change in inequality when wives' earnings are included as a source of income, i.e., (CVfamily—CVhead)/CVhead. The key components of this change are the share of total earnings attributable to wives' earnings relative to husbands', the correlation of spouses' earnings, and the dispersion of wives' and husbands' earnings.	Mitigating effect of wives' earnings actually increased slightly in all countries; the correlation of spouses' earnings would have to experience an unprecedented increase in order for wives' earnings to become disequalizing.
Corluy and Vandenbroucke (2013)	EU, 1995–2008.	ELFS and SILC; aged 20–59.	Decomposes household employment rate by individual employment rate, household structure, and jobs distribution over households. Decompose changes in at-risk-of-poverty rates on the basis of changes in the poverty risks of jobless household, and of other (nonjobless) households, and of changes in household joblessness due to individual employment rates, household structures and distribution of employment.	Incorrect to attribute disappointing poverty trends during the EU employment boom years solely to the modest conversion of individual employment successes in household employment successes, or more specifically to ongoing polarization of jobs over households. Complementarity of employment creation and poverty reduction through social transfers and inclusive labor market policies.

Continued

Table 18.A7 Household incomes and Earnings (see Section 18.2)—cont'd

Authors	Years and countries	Data sets* and sample selection	Method and important variables	Main findings
Daly and Valetta (2006)	United States; 1967–1989/ 1989–1998.	March CPS Demographic Supplement; including men with earnings equal to zero to account for the possibility that declining labor force participation by low-wage men contributed to rising inequality in family income. Equivalent family income	Semiparametric density estimation. For complete decomposition, four factors are considered: (i) distribution of men's earnings; (ii) women's labor force participation; (iii) family structure and (iv) underlying family characteristics, in this and by way of sensitivity test also in reversed order. The latter led to a somewhat larger role for residuals.	For the period 1969–1989, the growing dispersion of men's earnings and changing family structure can account for most of the rise in family income inequality. By contrast, the increase in labor force participation by women tended to offset this trend. Inequality grew at a slower rate in the 1990s largely because of stabilization in the relative earnings of men from low-income families. Larger effect found than by Burtless (1999) because of accounting for increasing inactivity. Consistent with "episodic" inequality change (Atkinson, 1997).
Del Boca and Pasqua (2003)	ITA; 1977–1998.	SHIW (and ECHP).	Decomposition of the CV2 of total household income. Three sources of income are considered: husband's earnings, wife's earnings and other sources of income (both from other components and nonlabor income). Simulations of household income distribution that would occur if wives had no earnings.	Total income distribution would have been more unequal without women's labor income.
Johnson and Wilkins (2003)	AUS; 1982–1997/ 1998.	Seven waves of the IDS.	Semiparametric procedure developed by DiNardo et al. (1996).	Changes in the distribution of work across families—for example, an increase in both two-earner families and no-earner families—were the single-most important source of the increase in private income inequality, with such changes on their own accounting for half the increase in inequality.

Karoly and Burtless (1995)	United States; 1959, 1969,1979, 1989.	Census and March CPS; Personal equivalent income distribution.	Decompose changes in Gini coefficient following Lerman and Yitzhaki (1984).	Increase in proportion of single-head families boosted inequality over entire period. Forty percent reduction in income inequality in the 1960s because of the decline in earnings inequality among male heads of families; more than one-third of increase in inequality after 1969 because inequality in male earnings soared. Since 1979 females' gains in earnings have increased inequality because these gains have been concentrated increasingly in families with high incomes.
Larrimore (2013)	United States; 1979–2007	March CPS, <1992 adjusted upward Square root equivalent income.	Shift share decomposition given Cowell-Fiorio's (2011) critique of Dinardo et al. (1996) and Daly and Valetta (2006) decomposition; the data intensity prevents this method from being suitable for all decompositions of interest. In particular, it is limited in its ability to observe how a range of income sources interact to account for changing inequality.	Factors contributing to rapid rise in income inequality in the 1980s differ substantially from those contributing to slower increase since that time. In the 1980s changes in the correlation of spouses' earnings accounted for income inequality growth, but not thereafter. Additionally, the 2000s business cycle is the first full business cycle in at least 30 years where changes in earnings of male household heads accounted for declines in income inequality. Instead, continued growth in income inequality was accounted for primarily by increases in female earnings inequality and declines in both male and female employment.
Lehrer (2000)	United States; 1973, 1992/1993.	NSFH; married couples.	Decompose CV2 of husband's plus wife's earnings.	1973 and 1992–1994 important similarity: spouse's contribution is equalizing in all life-cycle stages (no children, young <6 years, and older children). However, equalizing influence of wife's contribution grew substantially stronger—partly due to a decrease in the dispersion of female earnings relative to that of male earnings.

Continued

Table 18.A7 Household incomes and Earnings (see Section 18.2)—cont'd

Authors	Years and countries	Data sets* and sample selection	Method and important variables	Main findings
				Actual gap between "rich" and "poor" married-couple households, as measured by their income from labor, is narrower than if all wives were out of the labor force.
Lu et al. (2011)	CAN; 1980, 1995, 2005.	Census; heads 16–64; Census family as <2000 (opposite sex); excl. no earnings; square root equivalent income; head wages >0; full-time ≥30 h.	Semi-parametric decomposition methods Dinardo et al. (1996), closely following the work of Fortin and Schirle (2006). by male and female earnings structure, female EPOP, assortative mating, family compos. and characteristics.	1980–1995 substantial increases in family earnings inequality, some decrease 1995–2005 although earnings of Top 1% of families increase substantially. Employment rates of men and women, increases in their educational attainment, and decreases in assortative mating had equalizing effects (women coupling below their level); increases in the returns to higher education and in proportion of single individuals and lone-parent families drove increases in family earnings inequality.
Review of Economic Dynamics Special issue 2010: "Cross-sectional economic facts for macro-economists"	Relevant countries: CAN, DEU, ESP, ITA, SWE, UK, United States; 1960s/1970s/1980s/1990s up to mid-2000s.	National data sets on earnings, incomes and expenditures.	Country contributions by Brzozowski et al., Fuchs-Schündeln et al., Pijoan-Mas and Sanchéz-Marcos, Japelli and Pistaferri, Domeij and Floden., Blundell and Etheridge, and Heathcote et al. These document level and evolution, over time and over life cycle, of inequality of wages, labor earnings, income, consumption, and wealth, adopting as much as possible a uniform approach.	Substantial increases in wages and earnings inequality, over the last three decades; experience premium rose and gender premium fell virtually everywhere. Earnings inequality appears to be strongly counter-cyclical. In all countries, government redistribution through taxes and transfers reduced level, trend and cyclical fluctuations in income inequality. The rise in income inequality was stronger at the bottom of the distribution. Consumption inequality increased less than disposable income inequality, and tracked the latter much more closely at the top than at the bottom of the distribution.

				Measuring the age profile of inequality is challenging because of the interplay of time and cohort effects.
Shorrocks (1983)	United States; 1968–1977.	PSID, households excl. those with change of head.	Empirical approach to decomposition rules proposed by Shorrocks (1982).	Dollar for dollar, capital income and taxes have more distributional impact than earnings, which in turn exceeds the impact of transfer income (defined to include retirement pensions and annuities).
Van Weeren and Van Praag (1983)	BEL, DEU-W, DNK, FRA, ITA, NLD, UK; 1979.	Special survey (van Praag et al., 1982); net household income.	Between-group decomposition of variance of log incomes and Theil index by several socioeconomic characteristics.	In most countries the greatest inequality exists between employment subgroups (employees, self-employed, and not-working). Other important characteristics are age and education of the main breadwinner. The place of living household appeared to be of minor importance, and number of breadwinners is only of secondary importance.

Table 18.A8 Wage dispersion and the recent polarization and offshorability approaches to supply and demand (see Section 18.4.5)

Authors	Years and countries	Data sets* and sample selection	Method and important variables	Main findings
Acemoglu and Autor (2011)	Mainly United States, going back to around 1960, and 10 EU15 countries, going back to 1992.	Appendix provides detail of US data sources used for depicting trends; Truncate at bottom and top 5% of earnings distribution. Census is used for empirical example.	Takes stock of US trends in wage inequality in detail and adds some detail of polarization in EU countries. Evaluates shortcomings of canonical model explaining those and develops a model with endogenous assignment of three levels of skills to a continuum of tasks and possible substitution of machines for certain tasks previously performed by labor.	Provides a stylized empirical application of the new framework to US data, and suggests further directions for empirical exploration.
Antonczyk et al. (2010)	DEU-W, United States; 1979–2004.	IABS and ORG CPS, full-time working men.	This paper compares trends in wage inequality in the United States and Germany separating age, cohort, and time macro-economic effects. It accounts for potential cohort effects, an issue which is mostly ignored by the recent literature on wage inequality, even though SBTC may have a bias in the age/cohort dimension.	Between 1979 and 2004, wage inequality increased strongly in both the United States and Germany, but there were various country specific aspects of this increase. There is a large role played by cohort effects in Germany, whereas it is only small in the United States. Although there is evidence in both the United States and Germany, which is consistent with a technology-driven polarization of the labor market, the patterns of trends in wage inequality differ strongly enough that technology effects alone cannot explain the empirical findings. Episodic changes resulting from changes in institutional factors such as unionization or the minimum wage may explain the differences.
Autor (2013)	United States	Recent literature.	An emerging literature argues that changes in the allocation of workplace "tasks" between capital and labor, and between domestic and foreign workers, has altered the structure of labor demand in industrialized countries and fostered	The paper concludes with a cautiously optimistic forecast for the potential of the task approach to illuminate the interactions among skill supplies, technological capabilities, and trade and offshoring opportunities, in shaping the aggregate

			employment polarization—that is, rising employment in the highest and lowest paid occupations. Analyzing this phenomenon within the canonical production function framework is challenging, however, because the assignment of tasks is essentially static. This essay sketches an alternative model of the assignment of skills to tasks based on comparative advantage, reviews key conceptual and practical challenges that researchers face in bringing the "task approach" to the data, and cautions against two common pitfalls that pervade the growing task literature.	demand for skills, the assignment of skills to tasks, and the evolution of wages. For further research the classification of tasks is a challenge as the four task attributes—routine, abstract, manual, offshorable—though broadly distinct show important overlaps which hinder the classification of tasks. It is advisable to use, reuse, recycle, replicate, repeatedly apply existing task classifications, and thus attempt to converge upon a shared and standardized set of task measures. It is mistaken to give up on "middle-skill" education because there is no future for middle-skill jobs, as education is cumulative and middle-skill jobs are not slated to disappear though many middle-skill tasks may.
Autor and Dorn (2013)	United States; 1980–2005.	Census IPUMS and ACS.	The paper offers a unified analysis of the growth of low-skill service occupations and the concurrent polarization of employment and wages. It hypothesizes that polarization stems from the interaction between consumer preferences, which favor variety over specialization, and the falling cost of automating routine, codifiable job tasks. Applying a spatial equilibrium model where local labor markets have differential degrees of specialization in routine-intensive industries, it corroborates four implications of this hypothesis. Local labor markets that specialized in routine tasks differentially adopted information technology, reallocated low-skill labor into service occupations (employment polarization), experienced earnings growth at the tails of the distribution (wage polarization), and received inflows of skilled labor.	The twisting of the lower tail of the employment and earnings distributions is substantially accounted for by rising employment and wages in a single broad category. The paper considers a panoply of alternative explanations including offshoring of jobs tasks, income and substitution effects in high-skill consumption and labor supply, and demographic and economic shifts including immigration, population aging, female labor force entry, and declining manufacturing employment. Many of these alternative explanations receive some empirical support but none appears to play a leading role.

Continued

Table 18.A8 Wage dispersion and the recent polarization and offshorability approaches to supply and demand (see Section 18.4.5)—cont'd

Authors	Years and countries	Data sets* and sample selection	Method and important variables	Main findings
Autor et al. (2003)	United States; 1960–1998.	DOT occupational characteristics appended to Census IPUMS 1960–1990 and ORG CPS 1980–1998; employees aged 18–64, FTE weights.	The paper argues that computer capital (1) substitutes for workers in performing cognitive and manual tasks that can be accomplished by following explicit rules; and (2) complements workers in performing nonroutine problem solving and complex communications tasks. Provided that these tasks are imperfect substitutes, the model implies measurable changes in the composition of job tasks.	Within industries, occupations, and education groups, computerization is found to be associated with reduced labor input of routine manual and routine cognitive tasks and increased labor input of nonroutine cognitive tasks. Translating task shifts into education demand, the model can explain 60% of the estimated relative demand shift favoring college labor during 1970–1998. Task changes within nominally identical occupations account for almost half of this impact.
Blinder (2007)	United States; 2004.	O*NET.	Using detailed information on the nature of work done in over 800 BLS occupational codes, this paper ranks those occupations according to how easy/hard it is to offshore the work—either physically or electronically.	Using that ranking, it estimates that somewhere between 22% and 29% of all US jobs are or will be potentially offshorable within a decade or two. Because the rankings are subjective, two alternatives are presented—one objective, the other is an independent subjective ranking. It is found that there is little or no correlation between an occupation's "offshorability" and the skill level of its workers (as measured either by educational attainment or wages). However, it appears that, controlling for education, the most highly offshorable occupations were already paying significantly lower wages in 2004.
Blinder and Krueger (2009)	United States; 2008.	Special survey for Princeton Data Improvement Initiative (PDII).	This paper reports on a pilot study of the use of conventional household survey methods to measure something unconventional: what we call	Offshorability appears to be particularly prevalent in production work and in office and administrative jobs. By industry group, it is most common in manufacturing,

			"offshorability," defined as the ability to perform one's work duties (for the same employer and customers) from abroad. Notice that offshorability is a characteristic of a person's job, not of the person.	finance and insurance, information services, and professional and technical services. More educated workers appear to hold somewhat more offshorable jobs. But differences in offshorability by race, sex, age, and geographic region are all minor. In estimated multivariate econometric models, offshorability does not appear to have consistent systematic effects on either wages or the probability of layoff. union members and people in licensed positions are always less likely to hold offshorable jobs; and, perhaps surprisingly, routine work is no more likely to be offshorable than other work.
Dunne et al. (2004)	United States; 1977 and 1992.	March CPS and LRD; only plants that reported investments.	Using establishment-level data, we shed light on the sources of the changes in the structure of production, wages, and employment that have occurred over recent decades.	The findings are (1) the between-plant component of wage dispersion is an important and growing part of total wage dispersion; (2) much of the between-plant increase in wage dispersion is within industries; (3) the between-plant measures of wage and productivity dispersion have increased substantially over recent decades; and (4) a significant fraction of the rising dispersion in wages and productivity is accounted for by changes in the distribution of computer investment across plants.
Goos and Manning (2007) (see also 2003)	UK; 1976–1999.	NES complemented by for part-time workers.	The more nuanced version of SBTC recently proposed by Autor et al. (2003) makes a different prediction about what is happening to employment in low-wage jobs.	This paper presents evidence that employment in the United Kingdom is polarizing into lovely and lousy jobs in consistence with the nuanced view. Job polarization can explain one-third of the rise in the log(50/10) wage differential and one-half of the rise in the log(90/50).

Continued

Table 18.A8 Wage dispersion and the recent polarization and offshorability approaches to supply and demand (see Section 18.4.5)—cont'd

Authors	Years and countries	Data sets* and sample selection	Method and important variables	Main findings
Goos et al. (2011) (see also 2009 and 2010)	15 EU countries; 1996–2006.	LFS-EU, excl. agriculture and fishing, and OECD STAN	This paper develops a simple and empirically tractable model of labor demand to explain recent changes in the occupational structure of employment as a result of technology, offshoring, and institutions. This framework takes account not just of direct effects but indirect effects through induced shifts in demand for different products.	The routinization hypothesis of Autor et al. (2003) is found to be the most important factor behind the observed shifts in employment but offshoring does also play a role. Shifts in product demand are acting to attenuate the impacts of recent technological progress and offshoring. By implication, wage-setting institutions play little role in explaining job polarization in Europe.
Liu and Grusky (2013)	United States; 1979–2010.	ORG CPS, O*NET; nonmilitary wage and salary workers including part-time, aged 16–65.	Is the third industrial revolution indeed driven by rising payoffs to skill? This simple but important question has gone unanswered because conventional models of earnings inequality are based on exceedingly weak measurements of skill. By attaching occupational skill measurements to the CPS, it becomes possible to adjudicate competing accounts of the changing returns to cognitive, creative, technical, and social skill.	The well-known increase in between-occupation inequality is fully explained when such skills are taken into account, while returns to schooling prove to be quite stable once correlated changes in workplace skills are parsed out. The most important trend, however, is a precipitous increase in the wage payoff to synthesis, critical thinking, and related "analytic skills." The payoff to technical and creative skills, often touted in discussions of the third industrial revolution, is shown to be less substantial.
Mishel et al. (2013)	United States; 1973–2007.	ORG and May CPS (provide an independent test of earlier results based primarily on the decennial census and the American Community Survey);	The influential "skill-biased technological change" (SBTC) explanation claims that technology raises demand for educated workers, thus allowing them to command higher wages—which, in turn, increases wage inequality. A more recent SBTC explanation focuses on computerization's role in increasing employment in both higher-wage and lower-wage occupations,	Principal findings include: 1. Technological and skill deficiency explanations of wage inequality have failed to explain key wage patterns over the last three decades, including the 2000s. 2. History shows that middle-wage occupations have shrunk and higher-wage occupations have expanded since the 1950s. This has not driven any changed

		wage and salary workers aged 18–64.	resulting in “job polarization.” This paper contends that current SBTC models—such as the education-focused “canonical model” and the more recent “tasks framework” or “job polarization” approach mentioned above—do not adequately account for key wage patterns (namely, rising wage inequality) over the last three decades.	pattern of wage trends. 3. Evidence for job polarization is weak. 4. There was no occupational job polarization in the 2000s. 5. Occupational employment trends do not drive wage patterns or wage inequality. 6. Occupations have become less, not more, important determinants of wage patterns. 7. An expanded demand for low-wage service occupations is not a key driver of wage trends. 8. Occupational employment trends provide only limited insights into the main dynamics of the labor market, particularly wage trends.
Spitz-Oener (2006)	DEU-W 1979, 1985/1986, 1991/1992, 1998/1999.	Qualification and Career Survey BBIB; employees living in West Germany, German nationals, aged 18–65.	A unique data set from West Germany enables looking at how skill requirements have changed within occupations. Two hypotheses are tested: (1) IT is a substitute for routine manual and routine cognitive activities, and (2) IT is complementary to analytic and interactive activities.	Occupations are found to require more complex skills today than in 1979, and the changes in skill requirements have been most pronounced in rapidly computerizing occupations. It occurred within occupations, within occupation-education groups, and within occupation-age groups, changes in skill requirements similar to those in the United States. The question that now arises is why similar changes in skill requirements in all of these countries have not led to similar changes in the structure of wages.

Table 18.A9 Wage dispersion and institutions (see Section 18.4.6)

Authors	Countries and years	Data sets* and sample selection	Methods and important variables	Types of institutions**	Main findings
i. Overviews of the literature					
Blau and Kahn (1999)	13 EU countries and CHE, AUS, CAN, JPN, NZL, United States; 1970s to 1990s.	Draws on existing literature and data sources used there.	This chapter examines the impact of wage-setting institutions and government policies on wages and employment, focusing on the OECD countries.	AP, CP, DI, MW, UB, UD	There is considerable evidence that centralized collective bargaining, minimum wages, and antidiscrimination policies raise the relative wages of the low paid. Evidence of the impact of these institutions and other policies such as mandated severance pay, advance notice or unemployment insurance is more mixed with some studies finding active employment effects while others do not. This may reflect the adoption by many OECD countries of offsetting policies, such as public employment, temporary employment contracts and active labor market programs, which, although they may have reduced the adverse relative employment effects of their less-flexible labor market institutions on the low skilled, appear not to have prevented high overall unemployment.
Blau and Kahn (2009)	IALS: 8 countries; OECD: 12 countries.	IALS, 1994, 200+ annual hours and 10+ annual weeks; OECD Earnings database, 1980, 1990, 2000.	Documents and provides explanations for levels of and trends in earnings inequality, focusing on international (OECD) differences. Distinguishes between wage rates, hours worked, and earnings.	ED, HR, MW, UB	International differences reflect diversity of working population and prices, which in turn are affected by supply and demand as well as institutions. Collective bargaining and the minimum wage bring up the bottom, leading to employment losses. Offshoring deserves further attention (and may actually narrow wage differentials); so do employment protection, product market regulation and norms.

Doucouliagos and Stanley (2009)	United States	Extensive literature search.	Multivariate meta-regression analysis accommodate a potentially complex employment effect, misspecification biases and differential propensities to report adverse employment effects. It uses employment elasticity with respect to the minimum wage as the metric.	MW	Recently developed meta-analysis methods applied to 64 US minimum-wage studies (almost 1500 estimates) show that the minimum-wage effects literature is contaminated by publication selection bias, which is estimated to be slightly larger than the average reported minimum wage effect. Once this is corrected, little or no evidence of a negative association between minimum wages and employment remains. The results confirm those of the meta-analysis of Card and Krueger (1995b).
De Linde Leonard et al. (2013)	UK	Extensive literature search.	Multivariate meta-regression analysis of 236 estimated minimum-wage elasticities and 710 partial correlation coefficients from 16 UK studies.	MW	The study finds no overall practically significant adverse employment effect. Unlike US studies (see Doucouliagos and Stanley, 2009), there seems to be little, if any, overall reporting bias. It identifies several research dimensions that are associated with differential employment effects. In particular, the residential home care industry may exhibit a genuinely adverse employment effect.
Freeman (2005)	OECD	No new empirics.	This paper argues that there are two reasons for inconclusive debate over the claim that labor institutions impair aggregate performance. The first reason is that many adherents to the claim hold strong priors that labor markets operate nearly perfectly in the absence of institutions and let their priors dictate their modeling choices and interpretation of empirical results. The second reason is that the cross-country aggregate data at issue is weak—too weak to decisively reject strong prior views or to convince those with weaker priors.	Various	The debate over the influence of labor market flexibility on performance is unlikely to be settled by additional studies using aggregate data and making cross-country comparisons. Although this approach holds little promise, microanalysis of workers and firms and increased use of experimental methods represent a path forward. Steps along this path could help end the current “lawyer’s case” empiricism in which priors dominate evidence.

Continued

Table 18.A9 Wage dispersion and institutions (see Section 18.4.6)—cont'd

Authors	Countries and years	Data sets* and sample selection	Methods and important variables	Types of institutions**	Main findings
Freeman (2007)	OECD	No new empirics.	The paper documents the large cross-country differences in labor institutions that make them a candidate explanatory factor for the divergent economic performance of countries and reviews what economists have learned about the effects of these institutions on economic outcomes. It identifies three ways in which institutions affect economic performance: by altering incentives, by facilitating efficient bargaining, and by increasing information, communication, and trust.	CP, UB	The evidence shows that labor institutions reduce the dispersion of earnings and income inequality, which alters incentives, but finds equivocal effects on other aggregate outcomes, such as employment and unemployment. Given weaknesses in the cross-country data on which most studies focus, the paper argues for increased use of microdata, simulations, and experiments to illuminate how labor institutions operate and affect outcomes.
Katz and Autor (1999)	United States	No new empirics.	The chapter presents a framework for understanding changes in the wage structure and overall earnings inequality. It emphasizes the role of supply-and-demand factors and the interaction of market forces and labor market institutions. Recent changes in the US wage structure are analyzed in detail to highlight crucial measurement issues that arise in studying wage structure changes and to illustrate the operation of the framework. The roles of skill-biased technological	IR, MW, UB, UD	Several directions for future research are suggested: the roles of changes in labor market institutions (the incidence of labor market rents) and changes in competitive supply-and-demand factors. A key issue model is how to model the effects of institutions on employment rates and composition as well as on wages. The extent to which institutional changes reflect exogenous political events as opposed to responses to market forces can help sort out the effects of institutions from supply-and-demand factors. Taking a longer-term historical perspective will also be helpful as the US experience for

			change, globalization forces, changes in demographics and relative skill supplies, industry labor rents, unions, and the minimum wage in the evolution of the US wage structure arc examined, as are differences and similarities in wage structure changes among OECD nations.		the 1970s, 1980s, and 1990s illustrates. Cross-country comparative work and differences across regions within a country may also provide useful variation in demand-and-supply shocks and institutional factors.
Kierzenkowski and Koske (2012)	OECD	Recent literature.	Despite a general trend of increasing labor income inequality, there have been differences in the timing, intensity, and even direction of these changes across OECD countries. These stylized facts have led to numerous studies about the main determinants of labor income inequality and, as a result, a significant revision of the previous consensus about the key drivers. The most researched channels include skill-biased technological change, international trade, immigration, education, as well as the role of labor market policies and institutions.	MW, UB	SBTC (canonical view) fails to explain why inequality has diminished at the bottom relative to the median since the late 1980s as well as why within-group wage dispersion has grown substantially and mainly for college-educated workers. SBTC (nuanced view) explains why OECD labor markets have become polarized. International trade seems to have important implications for at least some groups of workers. Immigration has a rather small impact on native workers and sizable adverse wage or employment effects on the cohorts of previous immigrants. Education: Wage inequality is negatively correlated with the average level of educational attainment. Labor market policies and institutions: The impact of declining unionisation and of the lower relative minimum wage is most pronounced at the lower end of the wage distribution while cross-country evidence suggests that government employment reduces wage inequality.
Lemieux (2011)	AUS, CAN, United States; since 1980s.	No new empirics.	Wage inequality has been increasing in most industrialized countries over the last three decades. There are, nonetheless, major differences across	MW, UD, UW	Although demand factors linked to technological change may be a leading factor behind the secular growth in wage inequality (or the more recent polarization of wages), they cannot

Continued

Table 18.A9 Wage dispersion and institutions (see Section 18.4.6)—cont'd

Authors	Countries and years	Data sets* and sample selection	Methods and important variables	Types of institutions**	Main findings
			countries in terms of the timing and magnitude of the growth in inequality. A large number of explanations have been suggested for these observed changes, including technological progress and the computer revolution, labor market institutions and social norms, and changes in the relative supply of highly educated workers. This paper assesses the validity of these explanations in light of the large differences in inequality growth across countries, and the stunning growth in the concentration of income at the top end of the distribution.		account for the large differences in inequality growth observed across countries. Supply factors and institutions are more successful than demand at explaining differences across countries. None provides a compelling answer to the question of why inequality at the very top end of the distribution has increased so much in some countries but not in others. Two main conclusions are, first, that the SDI explanation is still alive and well, in the sense that no single explanation (supply, demand, or institution) can account for all of the changes in wage inequality observed across countries. Second, we still do not understand very well why inequality at the very top end of the distribution has increased so much in countries like the United States, Canada, or Australia.
Machin (1997)	UK; 1979 and 1993.		At the same time as the role of labor market institutions declined very dramatically in Britain there was a very sharp rise in wage inequality. It therefore provides a very good testing ground for evaluating the importance of labor market institutions in explaining the evolution of the wage structure. Regression of wage inequality measures distinguished by Wage Council applicability.	MW, UD	The weakening of unions and minimum wages that have traditionally propped up wage levels at the bottom end of the wage distribution, is found to play an important part in the rise in wage inequality in Britain.

Machin (2008)	Various.	No new empirics.	The paper describes the origins of the recent work documenting trends in wage inequality, the sizable body of research trying to understand national and international differences, and discuss the directions in which more recent work has moved and where it may go in future.	Various	It is concluded that the evidence shows that the wage distribution has been characterized by long-run growth in the relative demand for skills driven by technological change (rather than trade) and that changes in skill supply and institutional changes have affected the timing of how SBTC impacts on the wage structure in different contexts. Slower inequality growth in the lower tail in the United States and the United Kingdom and rising inequality in previously stable (European) distributions together with a polarization of job growth have refined explanations and added sophistication to the SBTC story.
Manning (2011)	International.	No new empirics.	The paper defends the claim that it is simply not true to claim that the perspective of perfect competition tells us all we need to know. There are rents in the typical jobs, though the size and distribution are not well known.	DI, IR, MW, UB, and employer collusion	Many empirical observations (e.g., equilibrium wage dispersion, the gender pay gap, geographical agglomeration, the effect of minimum wages on employment, employers paying for general training, costs of job loss for workers with no specific skills to list only a few) that are puzzles in the perspective of a perfectly competitive labor market are simply what one might expect if one thinks the labor market is characterized by pervasive imperfect competition. Views of the likely effects of labor market regulation should be substantially altered once one recognizes the existence of imperfect competition. However, although imperfect competition can be used as a justification for some regulation on efficiency grounds, it always predicts some limits to regulation with quite what those limits are left to empirical research to decide.

Continued

Table 18.A9 Wage dispersion and institutions (see Section 18.4.6)—cont'd

Authors	Countries and years	Data sets* and sample selection	Methods and important variables	Types of institutions**	Main findings
Nickell and Layard (1999)	Various subsets of OECD countries.	Various.	Aims to survey the literature to see propositions such as these depends on which labor market institutions really are bad for unemployment and growth, and which are not.	AP, CP, ED, MW, PM, TA, UB, UD, WE	There is quite strong evidence that the compressed earnings distributions in some OECD countries relative to the United States are a consequence of equally compressed skill distributions. Most of the gross features of unemployment and wage distributions across the OECD in recent years seem explicable by supply-and-demand shifts and the role required of special institutional features such as unions and minimum wages is correspondingly minimal. Labor market institutions on which policy should be focused are unions and Social Security systems. Encouraging product-market competition is a key policy to eliminate the negative effects of unions. For Social Security the key policies are benefit reform linked to active labor market policies to move people from welfare to work. By comparison, time spent worrying about strict labor market regulations, employment protection and minimum wages is probably time largely wasted.
Rogerson and Shimer (2011)	17 OECD countries; 1965 to late-2000s.	LFS-OECD and GGDC.	This chapter assesses how models with search frictions have shaped understanding of aggregate labor market outcomes in two contexts: business cycle fluctuations and long-run (trend)	Various	Results are mixed. Search models are useful for interpreting the behavior of some additional data series, but search frictions per se do not seem to improve our understanding of movements in total hours at either business-cycle frequencies

			changes (and the shock-and-institutions explanation of international differences). It consolidates data on aggregate labor market outcomes for a large set of OECD countries, and asks how models with search improve our understanding of these data.		or in the long run. Still, models with search seem promising as a framework for understanding how different wage-setting processes affect aggregate labor market outcomes.
Schmitt (2013)	United States	Recent literature since about 2000.	Labor markets have imperfections in the form of inadequate information, uneven bargaining power, limited ability to enforce long-term commitments, and insufficient insurance mechanisms against employment-related risks. Labor policies and institutions can in principle be used to address these imperfections. The report examines recent research on the employment effect of the minimum wage to determine the best current estimates of the impact of increases in the minimum wage on the employment prospects of low-wage workers. In particular, it discusses channels of adjustment to an increase in the minimum wage.	MW	The evidence points to little or no employment response to modest increases in the minimum wage. Evidence on 11 possible adjustments to minimum-wage increases may help to explain why the measured employment effects are so consistently small. The strongest evidence suggests that the most important channels of adjustment are: reductions in labor turnover; improvements in organizational efficiency; reductions in wages of higher earners ("wage compression"); and small price increases. These adjustment mechanisms appear to be more than sufficient to avoid employment losses, even for employers with a large share of low-wage workers.
World Bank (2012)	Worldwide	Recent literature since about 2000.	Labor markets have imperfections in the form of inadequate information, uneven bargaining power, limited ability	AP, CP, MW, TA, UB, WE	Policies should seek to avoid the distortive interventions that stifle labor reallocation and undermine the creation of jobs in functional cities and global value

Continued

Table 18.A9 Wage dispersion and institutions (see Section 18.4.6)—cont'd

Authors	Countries and years	Data sets* and sample selection	Methods and important variables	Types of institutions**	Main findings
			to enforce long-term commitments, and insufficient insurance mechanisms against employment-related risks. Labor policies and institutions can in principle be used to address these imperfections. It is important, then, to understand the role and the impacts of policies and institutions like labor market regulation, collective bargaining, active labor market programs, and social insurance. But the main constraints to the job creation often lie outside the labor market, and a clear approach is needed to support appropriate policy responses.		chains. But policies should also ensure voice and social protection, especially for the most vulnerable. Ideally, policies should aim at removing the market imperfections and institutional failures preventing the private sector from creating more of those jobs. If the constraints cannot be easily singled out or are difficult to remove, offsetting policies may be considered.
ii. Aggregate studies (institutions may be indirect)					
Acemoglu (2003)	AUS, BEL, CAN, DEU, DNK, FIN, ISR, NLD, NOR, SWE, UK, United States; mid-80s to mid-90s.	CPS and LIS; annual earnings of full-time, full-year male household heads aged of 18–64.	Relative-supply-demand model to determine differential effects, followed by a model of Differential Technology Responses across countries (not tested).	MW, UW, some Demographics	Relative demand for skills increased differentially across countries. Labor market institutions creating wage compression in Europe also encourage more investment in technologies increasing the productivity of less-skilled workers, implying less skill-biased technical change in Europe than the United States.
Alderson and Nielsen (2002)	16 OECD; 1967–1992.	Deininger and Squire (1996) for income inequality.	Gini of incomes regressed on economic aggregate measures.	OP, UD, WE	Direct investment and North–South trade have played a role in the determination of income inequality in the contemporary period; likewise for immigration.

Autor et al. (2005)	United States; 1973–2003.	March CPS and May/ORG CPS; real log hourly wages of wage and salary workers.	Extends quantile decomposition that nests DiNardo et al. (1996) in view of more differentiated developments over the earnings distribution after 1990s (as 90/50 and 50/10 trends diverge).	MW	Compositional shifts in labor force have contributed to earnings inequality during the 1990s.
Autor et al. (2008)	United States; 1963–2005.	March CPS, matched with DOT; FTFY workers log earnings. Abstract, routine and manual tasks within occupations.	Kernel reweighting approach of Lemieux used to facilitate a direct comparison. Overall 90/10, 90/50 and 50/10; between-group educational differentials; within-group 90/10, 90/50, and 50/10 residual wage gaps conditioned on measures of education, age/experience, and gender.	ED, occupations by mean years of schooling, MW	Upper-tail (90/50) inequality has increased steadily since 1980 and fluctuations in the real minimum wage are not a plausible explanation; a puzzling deceleration in relative demand growth for college workers in the early 1990s potentially reconciled by a modified version of the skill-biased technical change hypothesis that emphasizes the role of information technology in complementing abstract (high-education) tasks and substituting for routine (middle-education) tasks. Employment and wage growth by skill percentile are found to be positively correlated in each of the last two decades.
Blanchard and Wolfers (2000)	20 OECD; 1960–1995.	Nickell's and OECD institutions data.		AP, CP, ED, MW, PM, TA, UB, UD, WE	Interactions shock and institutions are essential for understanding international differences.
Bedard and Ferrall (2003)	AUS, BEL, CAN-BC/ON, DEU-W, FIN, FRA, JPN, NLD, SWE, UK, United States; 1964/1982 vs. 1969...1992	Wage data from national sources; test data from IME conducted in 1964 and 1982.	Compares test scores at age 13 to wages later in life.	ED	Wage dispersion, as summarized by Gini coefficients, is significantly related to test-score dispersion. For the United States, the United Kingdom, and Japan, with more data, evidence of skill-biased changes in wage dispersion between the early 1970s and the late 1980s is found.

Continued

Table 18.A9 Wage dispersion and institutions (see Section 18.4.6)—cont'd

Authors	Countries and years	Data sets* and sample selection	Methods and important variables	Types of institutions**	Main findings
Bertola and Boeri (2003)	EU15; 1982–1995.	OECD, Eurostat, ILO.	Institutions protective of labor serve some intended purpose. More intense competition may increase demand for protection, and certainly calls for reforms. A stylized model of the effects of structural change and resulting reform tensions is used to examine recent evidence.	CP, TA, UD, WE	Labor market reforms are becoming relatively more frequent in EMU countries, and many of them reduce welfare system generosity and deregulate labor markets. Most reforms are marginal, however, and in many cases deregulation-oriented reforms are accompanied by measures which appear to try and offset the implications of stronger competition instead. To exploit fully the advantages of economic and monetary integration, the institutional structure of labor and other markets needs to be revised extensively.
Budría and Pereira (2005)	DEU, FIN, FRA, GRC, ITA, NOR, PRT, SWE, UK; 1980s, 1990s.	EDWIN microdata, private-sector males ages 18–60, 35+ hours, nonagricultural employees.	Quantile regression and OLS of returns to education.	ED	Inequality increasing effect of tertiary education, through the "within" dimension, became more acute over last years.
Christopoulou et al. (2010)	AUT, BEL, DEU, ESP, GRC, HUN, IRL, ITA, NLD, 1995 and 2002.	EU-SES 1995, 2002; hourly wage including regular bonuses and payment for overtime.	Split between composition effect and returns effect, and residual.	MW, OP, UB, UD	Wage inequality growth diverges across countries. Only minor contribution of compositional change, but association with technology and globalization, while with immigration wages decline. Mixed effect of labor market institutions.
iii. Specific institutions in more depth (see iv for jobs polarization)					
Baccaro (2008)	51 Advanced, Central and Eastern European, Latin	New ILO data set on industrial relations and labor law, various	Between- and within-country regressions.	UB, UD, labor-law compliance	What changes from the 1990s on in advanced countries is the capacity of industrial relations institutions to reduce inequality directly by compressing market earnings. In particular, centralized

	American and Asian countries late 1980s to early 2000s; and analysis of 16 Advanced countries from the late 1970s.	dimensions of globalization, and controls for demand and supply of skilled labor.			collective bargaining seems to have become less redistributive than in the past. To the extent that industrial relations institutions continue to support and reproduce the welfare state, they reduce inequality indirectly though this channel.
Barth and Lucifora (2006)	AUT, BEL, DEU, DNK, ESP, FIN, FRA, GRC, ITA, NOR, SWE, UK; 1973...2003.	EDWIN, ECHP; nonagricultural employees 18–64 15+ hours, gross hourly earnings.	Model with supply and demand for different types of labor, as well as institutions affecting the bargained relative wage.	CP, ED, UB, UD	No evidence of increasing "over-education" in Europe. Bargaining coordination and employment protection have compressing effect on wages, but at different points of the wage distribution.
Bassanini and Duval (2006)	21 OECD; 1982–2003.		Aggregate employment and group-specific participation, institutional/policies interactions.	AP, MW, PM, TA, WE	Changes in policies and institutions appear to explain almost two-thirds of noncyclical unemployment changes over the past two decades.
Bassanini et al. (2009)	OECD; 1982–2003.	OECD annual cross-country aggregate data on the stringency of employment protection legislation and industry-level data on productivity.	Examines effect of dismissal regulation on productivity.	CP	Empirical results suggest that mandatory dismissal regulations have a depressing impact on productivity growth in industries where layoff restrictions are more likely to be binding. By contrast, no evidence is found of a productivity effect of regulations concerning temporary contracts, which suggests that partial reforms, facilitating the use of fixed-term and atypical contracts, are unlikely to have an important impact on efficiency and technological change and cannot therefore be a substitute for comprehensive reforms whereby dismissal restrictions for open-ended contracts are also weakened.

Continued

Table 18.A9 Wage dispersion and institutions (see Section 18.4.6)—cont'd

Authors	Countries and years	Data sets* and sample selection	Methods and important variables	Types of institutions**	Main findings
Bertola et al. (2001)	28 OECD countries; 1960–1999.	Database of Blanchard and Wolfers together with wage distributions from OECD Earnings database and labor force and population data from ILO.	Analysis why the United States moved from relatively high to relatively low unemployment over the last three decades. Institutions are largely assumed to be invariant.	AP, CP, TA, UB, UD, WE	Although macroeconomic and demographic shocks and changing labor market institutions explain a modest portion of this change, the interaction of these shocks and labor market institutions is the most important factor explaining the shift in the United States relative unemployment. This is consistent with Blanchard and Wolfers (2000). Controlling for country- and time-specific effects, high employment is associated with low wage levels and high levels of wage inequality. Disaggregating, the employment of both younger and older people fell sharply in other countries relative to the United States since the 1970s, with much smaller differences in outcomes among the prime-aged.
Bičáková (2006)	FRA, UK, United States; 1990–2002.	National LFS; 25–54; incl. self-employed; hourly wages (net in FR); skill groups: sex × age × education.	Focus is changes in the between-group variation in earnings, employment, unemployment, and inactivity; labor supply and demand model with heterogeneous types of labor, using a pseudo-panel of different skill-groups; three equations for wage, employment, and labor force participation as a function of exogenous supply and demand shifters, as implied by the structural model, is estimated by two-way fixed effects on group-level panel data.	Wage rigidity (MW in sideshow)	Trade-off inequality to unemployment for declining demand for low-skilled is found for FRA vs. UK and United States.

Bingley et al. (2013)	DNK; 1980–2003.	LFS; males aged 21–55 working full time in private sector.	Investigates the relationship between life-cycle wages and individual membership of unemployment insurance schemes, separating permanent from transitory wages and characterise them using membership of unemployment insurance funds.	WE	Unemployment insurance is associated with lower wage growth heterogeneity over the life cycle and greater wage instability, changing the nature of wage inequality from permanent to transitory. Robustness checks suggest that moral hazard is relevant.
Blau and Kahn (2002) (see also 1996)	OECD; 1979–1999	Various, both macro and micro.	Discusses the literature and builds on own earlier contributions to compare US labor market performance to other countries.	AP, CP, ED, TA, UB, WE	Interventionist labor market institutions in Europe compress wages and lower wage inequality; however, jobs most be lost for some groups. Institutional and demographic change and macroeconomic policy also differs to the advantage of the United States.
Boeri (2011)	Europe; since 1980.	fRDB-IZA social policy reforms database.	Reviews literature building on institutional reforms as quasi-natural experiments.	AP, CF, CP, TA, UB, WE	Literature is very informative but insufficiently accounts for long-lasting asymmetries between reformed and unreformed segments of the labor market. Extends Mortensen-Pissarides model with this segmentation for a theoretical approach that can help improving the identification of causal effects using reforms. Also gives empirical evidence on reforms.
Boeri and Jimeno (2005)	ITA; 1986–1995.	LFS rotation panel, Italian social security records (INPS archives).	Within-country exemptions to coverage of employment protection provisions allow making inferences on the impact of EPL when assessing the effects on dismissal probabilities and, using a change in EPL in 1990, on the equilibrium size distribution of firms.	CP	Results are in line with predictions of the theoretical model. Workers under permanent contracts in firms with less restrictive EPL are more likely to be dismissed. However, there is no effect on the growth of firms.

Continued

Table 18.A9 Wage dispersion and institutions (see Section 18.4.6)—cont'd

Authors	Countries and years	Data sets* and sample selection	Methods and important variables	Types of institutions**	Main findings
Bryson et al. (2012)	Europe and United States; early 2000s.	GSS 2002 and 2006, EWCS 2000 and 2005; employees with a permanent contract in private sector and in profit oriented firms only excluding managers and CEOs.	Presents new comparable data on the incidence of performance pay schemes. The percentage of employees exposed to incentive pay schemes ranges from around 10–15% in some European countries to over 40% in Scandinavian countries and the United States. Individual pay and profit/gain sharing schemes are widely diffused, whereas share ownership schemes are much less common, particularly in Europe.	UW	A number of empirical regularities are found. Incentive pay is less common in countries with a higher share of small firms. Higher product and labor market regulation are associated with lower use of incentive pay. Capital market development is a necessary requirement for a wider diffusion of incentive pay, particularly sharing and ownership schemes. Controlling for a large set of individual characteristics and company attributes, the probability that a worker is covered by an incentive scheme is higher in large firms and in high-skilled occupations, while it is much lower for females.
Card and DiNardo (2002)	United States; around 1970–2000.	CPS: March, May and ORG; diverging samples are compared.	Extended discussion of the measurement of technological change and of changes in the structure of wages in the U.S. labor market over the past 20–30 years, concluding to myriad shifts.	MW	Viewed from 2002, it now appears that the rise in wage inequality was an episodic event. A key problem for the SBTC hypothesis is that wage inequality stabilized in the 1990s despite continuing advances in computer technology; SBTC also fails to explain the evolution of other dimensions of wage inequality.
Card et al. (2004)	CAN, UK, United States; 1973/ 1984… 2001.	CAN-LFS + supplements, UK-LFS and GHS, May and ORG CPS; hourly wages of employees aged 16–65.	Comprehensive analysis of the evolution of unionization and wage inequality for both men and women in all three countries over the past two to three decades, as a sequel to Freeman (1980) and Freeman and Medoff (1984). The countries collect comparable data and share similar collective bargaining institutions.	UB, UD, MW	Unions reduce male inequality also after controlling for skill; but they increase female inequality; over time the declining unionization has eroded equalization.

Checchi and Garcia Peñalosa (2008)	11 EU countries, AUS, CAN, NOR, CHE and United States; 1969–2004.	LIS; na	Labor market institutions are a crucial determinant of wage inequality, the wage share in aggregate income, and the unemployment rate. Because these variables affect, in turn, the distribution of income across households, the question arises of whether stronger labor market institutions have an impact on income inequality. Institutions can in principle have conflicting effects. This paper examines what is the overall impact of labor market institutions on household income inequality. And counterfactually simulates adoption in other countries of labor standards of United States, UK or EU average.	CP, MW, TA, UB, UD	The evidence indicates that stronger institutions are associated with lower income inequality, but in some cases also with higher rates of unemployment. The magnitude of this trade-off is explored, and the changes in inequality and unemployment are quantified that would be observed if a common labor standard were imposed on members states of the European Union—results are not encouraging as a consequence of a lowering of employment protection; this could be accompanied though by a reinforcement of wage coordination and union density but these are no obvious policy targets.
Checchi and Garcia Peñalosa (2010)	OECD; 1960–2000.	Aggregate data from various sources (see paper's Appendix B).	This paper argues that personal income inequality depends on the wage differential, the labor share and the unemployment rate. Labor market institutions affect income inequality through these three channels, and their overall effect is theoretically ambiguous.	MW, TA, UB, UD	It is found that greater unionization and greater wage bargaining coordination have opposite effects on inequality, implying conflicting effects of greater union presence on income inequality.
Coelli et al. (1994)	14 OECD countries; around 1970–1990.	OECD National accounts sectoral data.	This paper examines the issue of wage flexibility in an international context using sectoral wage dispersion data from 14 OECD countries. It draws comparisons between a	UB	No strong systematic relationship exists between wage dispersion and the degree of centralization of labor market institutions.

Continued

Table 18.A9 Wage dispersion and institutions (see Section 18.4.6)—cont'd

Authors	Countries and years	Data sets* and sample selection	Methods and important variables	Types of institutions**	Main findings
			measure of wage dispersion and the degree of centralization of a country's wage setting institution to determine whether decentralized wage setting institutions are necessarily associated with more flexible wages. Inter-country comparisons are drawn among the levels of wage dispersion over time, and the relationship between wages and demand conditions for labor, including productivity and relative prices, are examined.		
Corsini (2008)	11 EU countries; early 1990s to early 2000s	BHPS, GSOEP and ECHP; employees.	The paper studies the evolution of wage differentials between graduate (skilled) and nongraduate (unskilled) workers. All countries show an increasing relative supply of skilled workers but different behaviors of the wage differentials. The standard explanation for nondecreasing differentials in the face of rising relative supply is that technological progress is skill biased. This in turn would imply that technological progress differs in its magnitude and effects across Europe. Turning	R&D and CP, UD, WE	The findings show that what is relevant in the determination of the differentials it is the pace and intensity at which technological progress takes place. Adding institutions to the role of R&D employment rates of different groups as well as union density and generosity of unemployment benefits are found to be important for explaining the evolution of the wage differentials between skilled and unskilled workers. They do not produce wage compression between skilled and unskilled workers.

			then to institutions a model is built of imperfect competition and wage bargaining which relates the differentials to the technological progress but also to several labor market institutions.		
Dahl et al. (2011)	DNK; 1992–2001.	IDA, Income Register; full-time workers aged 25–65 years employed in bargaining segments.	This paper studies how decentralization of wage bargaining from sector to firm-level influences wage levels and wage dispersion. We use detailed panel data covering a period of decentralization in the Danish labor market. The decentralization process provides variation in the individual worker's wage-setting system that facilitates identification of the effects of decentralization.	UB	We find a wage premium associated with firm-level bargaining relative to sector-level bargaining, and that the return to skills is higher under the more decentralized wage-setting systems. Using quantile regression, we also find that wages are more dispersed under firm-level bargaining compared to more centralized wage-setting systems.
Dinardo et al. (1996)	United States; 1979–1988.	CPS; hourly wages.	This paper presents a semiparametric procedure to analyze the effects of institutional and labor market factors on recent changes in the US distribution of wages. The effects of these factors are estimated by applying kernel density methods to appropriately weighted samples. The procedure provides a visually clear representation of where in the density of wages these various factors exert the greatest impact.	MW, UD	De-unionization and supply-and-demand shocks were important factors in explaining the rise in wage inequality from 1979 to 1988. The decline in the real value of the minimum wage explains a substantial proportion of this increase in wage inequality, particularly for women. Labor market institutions are as important as supply-and-demand considerations in explaining changes in the US distribution of wages from 1979 to 1988.
DiNardo and Lemieux (1997)	CAN, United States; 1981–1988.	CAN-LFS and CPS; men aged 17–64 excl. university	During the period 1981–1988 the decline in the percentage of workers belonging to unions and an increase in hourly wage	MW, UD	Results suggest that much more severe declines in the unionization rate in the United States than in Canada account for two-thirds of the differential growth in

Continued

Table 18.A9 Wage dispersion and institutions (see Section 18.4.6)—cont'd

Authors	Countries and years	Data sets* and sample selection	Methods and important variables	Types of institutions**	Main findings
		graduates 17–19.	inequality were much more pronounced in the United States than in Canada. Study the effect of labor market institutions on changes in wage inequality by computing simple counterfactuals such as the distribution of wages that would prevail if all workers were paid according to the observed nonunion wage schedule.		wage inequality between the two countries.
Dustmann et al. (2009)	DEU-W; mid-1970s to mid-2000s.	IABS 1975–2004, and LIAB 1995–2004; ages 21–60.	Using the kernel reweighting procedure (DiNardo et al., 1996) it is shown that it is important to account for changes in workforce composition, in particular at the upper end of the wage distribution. Fluctuations in relative supply explain the evolution of the wage differential between the low- and medium-skilled very well, but do a poor job in predicting the evolution of the wage differential between the medium- and high-skilled.	UD	Wage inequality in West Germany has increased over the past three decades, contrary to common perceptions. During the 1980s, the increase was concentrated at the top of the distribution; in the 1990s, it occurred at the bottom end as well. Technological change is responsible for the widening of the wage distribution at the top. At the bottom of the wage distribution, the increase in inequality is better explained by episodic events, such as supply shocks and changes in labor market institutions. Occupations with high median wages in 1980 experienced the highest growth rate, whereas occupations in the middle of the 1980 wage distribution lost ground relative to occupations at the bottom.
Eissa and Hoynes (2004)	United States; 1984–1996.	March CPS; married couples residing in the	Simulation of 1984 and 1996 EITC rules on married couples labor participation. Effects estimated using both quasi-	TA	EITC family targeting can disincentivize secondary earners: 1% fall married women, strong increase for single-parent women, slight increase married men.

		same household, ages 25–54, and less than high school in main estimates	experimental and traditional reduced-form labor supply models, with same conclusion.		
Firpo et al. (2011)	United States; 1976/1977, 1988/1990, 2000/2002, 2004/2004, 2009/2010.	CPS and O*Net; male employees	Changes in returns to occupational tasks have contributed to changes in the wage distribution over the last three decades. Using a decomposition based on Firpo et al. (2009).	OP, UD	Technological change and deunionization played a central role in the 1980s and 1990s, and offshorability became an important factor from the 1990s onward.
Fortin and Lemieux (1997)	United States; 1979 and 1988.	CPS; workers aged 16–65.	Show what the variance of the (log) wage distribution would have been, if each of the three institutional changes had not happened. Decompose distribution of wages using three elements: the fraction of workers "affected" by the institutional factor of interest; the mean level of log wages among affected and nonaffected workers; and the dispersion of log wages among affected and nonaffected workers. By reverting some of these measures to their previous level, simulate what would have happened if the institutional changes had not taken place.	MW, PM, UB, UD	Historical evidence from the United States, international comparisons among industrialized countries and analyses of US data for the 1980s all yield the same conclusion: institutional forces simply cannot be overlooked in any serious attempt to understand the recent rise in wage inequality in the US labor market.
Golden and Wallerstein (2006)	16 OECD countries; 1980–2000.		Examine three main hypotheses for the rise of pay inequality: postindustrial, globalization, and institutional. Main idea is	OP, UB, UD, WE	Causes for pay inequality are quite different in the 1980s than in the 1990s. In the 1980s, growing wage dispersion is due to changes in the institutions of the labor

Continued

Table 18.A9 Wage dispersion and institutions (see Section 18.4.6)—cont'd

Authors	Countries and years	Data sets* and sample selection	Methods and important variables	Types of institutions**	Main findings
			determinants of wage inequality underwent considerable substantive change over the period. A statistical model uses first differences over 5-year periods, because effects of the explanatory variables are not instantaneous.		market. Declining unionization and declines in the level at which wages are bargained collectively both contribute to widening pay dispersion in the 1980s. In the 1990s, by contrast, increases in pay inequality are due to increasing trade with less developed nations. To the extent that low-pay workers have been protected from rising wage differentials in the 1990s, it has been because of government policy, in the form of social insurance, and not thanks to labor organizations.
Hall and Krueger (2010)	United States; 2008.	Special survey of a representative sample of US workers to inquire about the wage determination process at the time they were hired into their current or most recent jobs.	Some workers bargain with prospective employers before accepting a job. Others face a posted wage as a take-it-or-leave-it opportunity. Theories of wage formation point to substantial differences in labor market equilibrium between bargained and posted wages. A third of the respondents reported bargaining over pay before accepting their current jobs. About a third of workers had precise information about pay when they first met with their employers, a sign of wage posting. About 40% of workers could have remained on their earlier jobs at the time they accepted their current jobs, indicating a more favorable bargaining position than is held by unemployed job-seekers.	UW	Our analysis of the distribution of wages shows that wage dispersion is higher among workers who bargained for their wages. Wages are higher among bargainers than nonbargainers, after adjusting for the differing compositions of the groups. Our results on wages give substantial support to the job-ladder model—workers who had the option to remain at their earlier jobs when they took their current jobs can earn higher wages than those without that option.

Kenworthy (2001)	AUS, AUT, BEL, CAN, CHE, DEU, DNK, FIN, FRA, ITA, JPN, NLD, NOR, SWE, UK, United States	15 Bargaining Indicators found in the literature.	This article offers a survey and assessment of the principal existing measures in the literature: eight measures of wage centralization and seven measures of wage coordination. There are three aims: provide an inventory of existing indicators, examine their features and merits, and assess sensitivity of findings generated by these measures.	UB	The two best available measures of centralization of wage bargaining are the Iversen and Traxler–Blaschke–Kittel indicators. The former is based on structural features, and the latter aims to measure behavior. There is currently only one available measure of wage-setting centralization. The conceptual differences between wage-setting measures lead to some noteworthy differences in scoring of certain countries and years. A potentially problematic gap is the lack of any measure of wage setting at the subnational level.
Koeniger et al. (2007)	AUS, CAN, FIN, FRA, DEU, ITA, JPN, NLD, SWE, UK, United States; 1970s, 1980s, 1990s.	Various.	Variance decomposition of aggregates.	C, MW, U, WE	Institutions explain at least as much as trade and technology.
Kugler and Pica (2008)	ITA; 1986–1995.	Social Security employer–employee panel.	Study effects of the Italian reform of 1990 on worker and job flows, exploiting the fact that this reform increased unjust dismissal costs for businesses below 15 employees, while leaving dismissal costs unchanged for bigger businesses, to set up a natural experiment research design.	CP	The increase in dismissal costs decreased accessions and separations for workers in small relative to large firms, especially in sectors with higher employment volatility, with a negligible impact on net employment. Also some evidence is found suggesting that the reform reduced firms' entry rates and employment adjustments, but had no effect on exit rates.
Lemieux (2008)	United States; 1973–2005.	May and ORG CPS, PSID; mainly males but females partly considered separately.	The paper reviews recent developments in the literature on wage inequality with a particular focus on why inequality growth has been particularly concentrated in the top end of the wage distribution over the	MW, NO, UB, UD, UW	The nature of the changes in inequality has been dramatically altered over the last 15 years. Although the growth in inequality in the 1980s was pervasive, it has been concentrated at the top end of the distribution since then unlike SBTC, the institutional change explanation can

Continued

Table 18.A9 Wage dispersion and institutions (see Section 18.4.6)—cont'd

Authors	Countries and years	Data sets* and sample selection	Methods and important variables	Types of institutions**	Main findings
			last 15 years. Several possible institutional and demand-side explanations are discussed for the secular growth in wage inequality in the United States and other advanced industrialized countries.		help explain why inequality changes became concentrated in the top end after 1990 and why inequality grew more in the United States and the United Kingdom than in other advanced countries. This being said, just like in the 1980s, available estimates indicate that institutional change can only account for about a third of the observed recent changes in wage inequality. However, broadening the traditional institutional explanation to include pay-setting mechanisms such as performance-pay can help explain more of the growth in inequality at the top end. For the time being, however, most of the growth in top-end inequality over the last 15 years remains unaccounted for.
Lemieux et al. (2009)	United States; 1976–1998.	PSID (some robustness test using NLSY); male household heads, aged 18–65, employees in private sector.	An increasing fraction of jobs explicitly pay workers for their performance using bonus pay, commissions, or piece-rate contracts. Variance components analysis.	UD, UW	Compensation in performance-pay jobs is more closely tied to both observed and unobserved productive characteristics of workers than compensation in nonperformance-pay jobs. The return to these productive characteristics increased faster over time in performance-pay jobs. Performance pay provides a channel through which underlying changes in returns to skill get translated into higher wage inequality, accounting for 21% of the growth in the variance of male wages between the late 1970s and the early 1990s and for most of the increase in wage inequality above the 80th percentile over the same period.

Leonardi and Pica (2013)	ITA; 1985–1997.	Italian Social Security Institute (INPS) matched employer–employee panel: Veneto Workers History data set; private-sector excluding agriculture, male employees aged 20–55.	This study estimates the effect of employment protection legislation on wages, exploiting the 1990 Italian reform that introduced unjust dismissal costs for firms below 15 employees. It combines a regression discontinuity design (RDD) with a difference-in-difference (DID) approach for identifying the effect.	CP	The slight average wage reduction induced by the reform hides highly heterogeneous effects. Workers who change firm during the reform period suffer a drop in the entry wage, while incumbent workers are left unaffected. Also, the negative effect of the reform is stronger for young blue collars, low-wage workers and workers in low-employment regions. This pattern suggests that the ability of employers to shift firing costs onto wages depends on workers' relative bargaining power.
Levy and Temin (2007)	United States; 1930s to mid-2000s.	Various.	We provide a comprehensive view of widening income inequality in the United States contrasting conditions since 1980 with those in earlier postwar years. We argue that the income distribution in each period was strongly shaped by a set of economic institutions. A Bargaining Power Index is used (percent of output captured by full-time worker's compensation), split by categories of workers.	MW, OP, TA, UB	The early postwar years were dominated by unions, a negotiating framework set in the Treaty of Detroit, progressive taxes, and a high minimum wage—all parts of a general government effort to broadly distribute the gains from growth. More recent years have been characterized by reversals in all these dimensions in an institutional pattern known as the Washington Consensus. Other explanations for income disparities including skill-biased technical change and international trade are seen as factors operating within this broader institutional story.
Manzo and Bruno (2014)	United States construction industry; 2007–2011.	IPUMS data from the ACS, 5.0% sample.	An ordinary least squares (OLS) regression model is run evaluating the effects on decile ratios of income inequality of unionization, distinguishing	UD, UB	The largest contributor to rising income inequality has been the gradual, long-term decline in labor union membership. The union wage premium is between 10 and 17%, helping lower- and

Continued

Table 18.A9 Wage dispersion and institutions (see Section 18.4.6)—cont'd

Authors	Countries and years	Data sets* and sample selection	Methods and important variables	Types of institutions**	Main findings
			between prevailing-wage-law states and right-to-work law states, and controlling for demographic, educational, and work factors, including 24 distinct occupations.		middle-income workers most. Right-to-work laws decrease unionization by between 5% and 8% points and reduce the average construction worker's earnings by 6% in the national economy.
Nunziata (2005)	OECD; 1960–1994.	Various.	An empirical analysis of the determinants of labor cost, with particular reference to the impact of labor market institutions from 1960 to 1994. The paper also discusses the econometric issues related to the estimation of a macro pooled model like ours: among other things, the hypothesis of poolability and the cointegration properties of the model. The explanatory power of the model is finally tested by means of a series of country by country dynamic simulations.	CP, MW, TA, UB, UD, UW, WE	Labor market regulations can explain a large part of the labor cost rise in the last few decades once we control for productivity.
Oliver (2008)	14 OECD countries; 1980–2002.	Unpublished data set from the OECD	With a series of cross-sectional time-series analyses, this article investigates how a particular wage-bargaining institution: the extent to which industrywide wage minima (wage scales) cover both higher and lower skilled workers, mitigates pressures from growing international competition and new production techniques and affects the degree of wage inequality growth.	UW	The results strongly indicate that the presence of industrywide wage scales is a key factor in the evolution of wage inequality across OECD countries.

Plotnick (1982)	United States; 1957–1977.	Unpublished earnings data Henle and Ryscavage (1980), CPS (unpublished and several specific data sources; males).	This study uses newly available time series data to analyze trends in earnings inequality. It shows that although a human capital approach fits the data well and most of its predictions on signs are correct, the model's more exacting implications are not satisfied. A complementary more ad hoc approach retains variables found to be significant and looks beyond aggregate inequality measures into parts of the distribution which gain or lose.	UD	The major finding is that the observed slow upward trend in earnings inequality is well explained by a small number of plausible economic factors. Earnings inequality is significantly related to the level of unionization, dispersion in weeks worked, the age distribution of workers, and inequality of education. Once such factors are considered, there was no secular trend in earnings inequality over the 1958–1977 period.
Scheve and Stasavage (2009)	AUS, CAN, CHE, DEU, DNK, FRA, IRL, JPN, NLD, NZL, SWE, UK, United States; 1916–2000.	Top-incomes data, OECD earnings database, Lydall (1968), existing data on political institutions and new political data coded by the authors.	Although explaining post-1970 differences in income inequality between OECD countries is an important task, it is also the case that convincing comparative political economy hypotheses should be able to account as well for inequality trends in earlier time periods. The article considers the correlation of centralized wage bargaining and government partisanship with three separate top incomes fractions. With a longer time span there has been significantly more variation within countries over time than there has been between countries. A longer time span also enables examining whether within-country changes in institutions like wage bargaining centralization have been associated with changes in inequality.	UB	Regression analysis over the 13 countries and the whole period gives little evidence that government partisanship and wage-bargaining centralization can account for variation in inequality over the long run. A test of four individual countries that established a centralized system of wage bargaining in the middle of the twentieth century (DNK, IRL, NLD, SWE) also shows little evidence of an effect on inequality. This raises questions about the extent to which centralized wage bargaining is an institution that has a causal effect on inequality or alternatively whether centralized bargaining is simply an outcome that has, along with income equality, evolved over time in response to an underlying political or economic process.

Continued

Table 18.A9 Wage dispersion and institutions (see Section 18.4.6)—cont'd

Authors	Countries and years	Data sets* and sample selection	Methods and important variables	Types of institutions**	Main findings
Schivardi and Torrini (2008)	ITA; 1986–1998.	INPS comprehensive longitudinal matched employer–employees data set.	The paper studies the effects of the more stringent employment protection legislation that applies to firms with over 15 employees. It considers firms' propensity to grow when close to that threshold and changes in employment policies when they pass it. Using a the stochastic transition matrix for firm size.	CP	The probability of firms' growth is reduced by around 2% points near the threshold. The long-run effects of EPL on the size distribution of firms are quantitatively modest. Contrary to the implications of more stringent firing restrictions, workers in firms just above the threshold have on average less stable employment relations than those just below it; this might be because firms above the threshold make greater use of flexible employment contracts, arguably to circumvent the stricter regulation on open-end contracts.
Van der Wiel (2010)	NLD; 1997–2001.	Dutch Socio-Economic Panel SEP1984–2002: five waves containing contractual information.	This paper empirically establishes the effect of the employer's term of notice on the wage level of employees through a fixed effects regression model. The term of notice is defined as the period an employer has to notify workers in advance of their upcoming dismissal. The wages paid during this period are an important element of firing costs and hence employment protection. To find a causal effect, the paper exploits the exogenous change in the term of notice that resulted from the introduction of a new Dutch law in 1999.	CP	Strong evidence is found that a longer "dormant" term of notice leads to higher wages. In the sample used, an additional month of notice increases wages by 3%, ceteris paribus.

*Data set abbreviations	**Codes for categories of institutions
ACS: American Community Survey (United States)	AP: active labor market policies (concerning unemployed)
BHPS: British Household Panel Survey (UK)	CF: contract—flexibility
CPS: Current Population Survey—ORG: Outgoing Rotation Groups (United States)	CP: contract—employment protections
Deininger & Squire: Income Inequality Data set (World Bank)	DI: discrimination (gender)
DOT: Dictionary of Occupational Titles (United States)	ED: educational system
ECHP: European Community Household Panel (EU, Eurostat)	HR: working hours (part-time)
EDWIN: Education and Wage Inequality in Europe (EU Research Project)	IR: industry (rents)
EU-SES: European Structure of Earnings Survey (EU, Eurostat)	MW: minimum wage
EWCS: European Working Conditions Surveys (EU, European Foundation)	NO: Norms
fRDB-IZA: Labor market institutions and labor market reforms database (International, at Fondazione Rodolfo Debenedetti and IZA)	PM: product market (deregulation, competition)
GGDC: Groningen Growth and Development Centre (International, at University of Groningen)	OP: openness (globalization, migration, offshorability)
GSS: General Social Survey (United States)	TA: taxation
GHS: General Household Survey (UK)	UD: union—density
GSOEP: German Socioeconomic Panel (DEU)	UB: union—bargaining coverage, bargaining coordination, collective labor agreements
HBAI: Households Below Average Income UK)	UW: union—wage outcomes, pay systems (performance pay)
HES: Household Expenditure Survey (AUS)	WE: welfare (UI, benefits, EITC, in-work benefits)
IABS: IAB Employment Samples (DEU)	
IALS: International Adult Literacy Survey (OECD/Statistics Canada)	
IDA: Longitudinal data of Integrated Database of Labor Market Research (DNK)	
IDS: Income Distribution Survey (AUS)	
IME: International Mathematics Examinations (International)	
IPUMS: Integrated Public Use Microdata Series: IPUMS-CPS and IPUMS-United States for Census/ACS (United States)	
LFS: Labour Force Survey (CAN, EU, OECD, and UK)	
LIAB: Linked Employer-Employee Data from the IAB (DEU)	
LIS: Luxembourg Income Study (International)	
LRD: Longitudinal Research Database (United States)	
NES: New Earnings Survey (UK)	
NLSY: National Longitudinal Survey of Youth (United States)	
NSFH: National Survey of Families and Households (United States)	
OECD STAN: Structural Analysis Database (International, at OECD)	
O*NET: Occupational Information Network (United States)	
PSID: Panel Study of Income Dynamics (United States)	
SHIW: Survey of Household Income and Wealth (ITA)	
SILC: Statistics on Income and Living Conditions (EU, Eurostat)	
Top-incomes data: World Top Incomes Database (International, at Paris School of Economics)	

REFERENCES

Abowd, John, Kramarz, Francis, Lemieux, Thomas, Margolis, David, 1999. Minimum wages and youth employment in France and the United States. In: Blanchflower, David, Freeman, Richard (Eds.), Youth Employment and Joblessness in Advanced Countries. NBER and University of Chicago Press, Boston/Chicago (Chapter 11).

Acemoglu, Daron, 2003. Cross-country inequality trends. Econ. J. 113, F121–F149.

Acemoglu, Daron, Autor, David, 2011. Skills, tasks and technologies: implications for employment and earning. In: Ashenfelter, Orley, Card, David (Eds.), Handbook of Labor Economics. vol. 4/B. Elsevier, Amsterdam, pp. 1043–1171.

Acemoglu, Daron, Autor, David, 2012. What Does Human Capital Do? A Review of Goldin and Katz's The Race Between Education and Technology, NBER Working Paper 17820.

Agell, Jonas, Lommerud, Kjell Erik, 1992. Union egalitarianism as income insurance. Economica 59 (235), 295–310.

Agell, Jonas, Lommerud, Kjell Erik, 1993. Egalitarianism and growth. Scand. J. Econ. 95 (4), 559–579.

Aghion, Philippe, Algan, Yann, Cahuc, Pierre, 2011. Civil society and the state: the interplay between cooperation and minimum wage regulation. J. Eur. Econ. Assoc. 9 (1), 3–42.

Alderson, Arthur S., Nielsen, François, 2002. Globalisation and the great U-turn: income inequality trends in 16 OECD countries. Am. J. Sociol. 107, 1244–1299.

Alesina, Alberto F., Algan, Yann, Cahuc, Pierre, Giuliano, Paola, 2010. Family Values and the Regulation of Labor, NBER Working Paper 15747.

Allegretto, Sylvia, Dube, Arindrajit, Reich, Michael, 2011. Do minimum wages really reduce teen employment? Accounting for heterogeneity and selectivity in state panel data. Indus. Relat. 50 (2), 205–240.

Alvaredo, Facundo, 2011. A note on the relationship between top income shares and the Gini coefficient. Econ. Lett. 110 (3), 274–277.

Alvaredo, Facundo, Atkinson, Anthony B., Piketty, Thomas, Saez, Emmanuel, 2013. The top 1 percent in international and historical perspective. J. Econ. Perspect. 27 (3), 3–20.

Amable, Bruno, 2003. The Diversity of Modern Capitalism. Oxford University Press, Oxford.

Andersson, Frederik, Freedman, Matthew, Haltiwanger, John, Lane, Julia, Shaw, Kathryn, 2006. Reaching for the Stars: Who Pays for Talent in Innovative Industries? NBER Working Paper 12435.

Antonczyk, Dirk, DeLeire, Thomas, Fitzenberger, Bernd, 2010. Polarization and Rising Wage Inequality: Comparing the U.S. and Germany, Discussion Paper 10-015. ZEW, Mannheim.

Ashenfelter, Orley, Card, David (Eds.), 2011. Handbook of Labor Economics, vol. 4/B. Elsevier, Amsterdam.

Atkinson, Anthony B., 1970. On the measurement of inequality. J. Econ. Theor. 2, 244–263.

Atkinson, Anthony B., 1993. What is happening to the distribution of income in the UK? Keynes lecture in economics 1992. Proc. Br. Acad. 82, 317–351, Oxford University Press.

Atkinson, Anthony B., 1997. Bringing income distribution in from the cold. Econ. J. 107 (441), 297–321.

Atkinson, Anthony B., 1999. Economic Consequences of Rolling Back the Welfare State. MIT Press and CESifo, Munich.

Atkinson, Anthony, 2007a. The long run earnings distribution in five countries: "remarkable stability," U, V or W? Rev. Income Wealth 53 (1), 1–25.

Atkinson, Anthony, 2007b. The distribution of earnings in OECD countries. Int. Labour Rev. 146 (1–2), 41–60.

Atkinson, Anthony B., 2008. The Changing Distribution of Earnings in OECD Countries. Oxford University Press, Oxford.

Atkinson, Anthony B., 2009. Factor shares: the principal problem of political economy? Oxf. Rev. Econ. Pol. 25 (1), 3–16.

Atkinson, Anthony B., Bourguignon, François, 2000. Introduction. In: Handbook of Income Distribution, vol. 1. Elsevier, North Holland.

Atkinson, Anthony, Brandolini, Andrea, 2001. Promise and pitfalls in the use of secondary datasets: income inequality in OECD countries as a case study. J. Econ. Lit. 39, 771–799.

Atkinson, Anthony B., Brandolini, Andrea, 2006. From earnings dispersion to income inequality. In: Farina, F., Savaglia, E. (Eds.), Inequality and Economic Integration. Routledge, London (Chapter 2).

Atkinson, Anthony B., Morelli, S., 2012. Chartbook of Economic Inequality: 25 Countries 1911–2010. http://ineteconomics.org/sites/inet.civicactions.net/files/Note-15-Atkinson-Morelli.pdf.

Atkinson, Anthony B., Piketty, Thomas (Eds.), 2007. Top Incomes Over the Twentieth Century: A Contrast Between Continental European and English-Speaking Countries. Oxford University Press, Oxford.

Atkinson, Anthony B., Piketty, Thomas (Eds.), 2010. Top Incomes. A Global Perspective. Oxford University Press, Oxford.

Atkinson, Anthony B., Rainwater, Lee, Smeeding, Timothy, 1995. Income Distribution in European Countries, Working Paper MU 9506, Department of Applied Economics, University of Cambridge.

Atkinson, Anthony B., Piketty, Thomas, Saez, Emmanuel, 2011. Top incomes in the long run of history. J. Econ. Lit. 49 (1), 3–71.

Autor, David, 2013. The "Task Approach" to Labor Markets: An Overview, NBER Working Paper 18711.

Autor, David, Dorn, David, 2013. The growth of low-skill service jobs and the polarization of the US labor market. Am. Econ. Rev. 103 (3), 1553–1597.

Autor, David, Levy, Frank, Murnane, Richard J., 2003. The skill content of recent technological change: an empirical exploration. Q. J. Econ. 116 (4), 1279–1333.

Autor, David, Katz, Lawrence, Kearney, Melissa, 2005. Rising Wage Inequality: The Role of Composition and Prices, NBER Working Paper 11628.

Autor, David, Katz, Lawrence F., Kearney, Melissa S., 2006. The polarization of the U.S. labor market. Am. Econ. Rev. Papers Proc. 96 (2), 189–194.

Autor, David, Katz, Lawrence, Kearney, Melissa, 2008. Trends in U.S. wage inequality: revising the revisionists. Rev. Econ. Stat. 90 (2), 300–323.

Autor, David, Manning, Alan, Smith, Christopher, 2010. The Contribution of the Minimum Wage to U.S. Wage Inequality Over Three Decades: A Reassessment, NBER Working Paper 16533.

Baccaro, Lucio, 2008. Labour, Globalisation and Inequality: Are Trade Unions Still Redistributive? International Institute for Labour Studies, Discussion Paper 192. ILO, Geneva.

Barth, E., Lucifora, C., 2006. Wage Dispersion, Markets and Institutions: The Effects of the Boom in Education on the Wage Structure, Discussion Paper 2181, IZA, Bonn.

Bassanini, Andrea, Duval, Romain, 2006. The determinants of unemployment across OECD countries: reassessing the role of policies and institutions. OECD Econ. Stud. 2006 (1), 7–86.

Bassanini, Andrea, Nunziata, Luca, Venn, Danielle, 2009. Job protection legislation and productivity growth in OECD countries. Econ. Policy 24 (4), 349–402.

Bauer, Thomas, Bender, Stephan, Bonin, Holger, 2007. Dismissal protection and worker flows in small establishments. Economica 74 (296), 804–821.

Becher, Michael, Pontusson, Jonas, 2011. Whose interests do unions represent? Unionization by income in western Europe. In: Brady, David (Ed.), Comparing European workers part B: policies and institutions. Res. Sociol. Work 22, Emerald, Bingley, pp. 181–211.

Bedard, Kelly, Ferrall, Christopher, 2003. Wage and test score dispersion: some international evidence. Econ. Educ. Rev. 22 (1), 31–43.

Behar, Alberto, 2013. The Endogenous Skill Bias of Technical Change and Inequality in Developing Countries, IMF Working Paper No. 13/50.

Bell, Linda A., Freeman, Richard B., 2001. The incentive for working hard: explaining hours worked differences in the US and Germany. Labour Econ. 8, 181–202.

Belot, Michèle, van Ours, Jan, 2004. Does the recent success of some OECD countries in lowering their unemployment rates lie in the clever design of their labour market reform? Oxford Econ. Papers 56 (4), 621–642.

Benassi, Chiara, 2011. The Implementation of Minimum Wage: Challenges and Creative Solutions, Global Labour University, Working Paper 12, ILO, Geneva.

Bertola, Giuseppe, 2004. A pure theory of job security and labour income risk. Rev. Econ. Stud. 71 (1), 43–61, Oxford University Press.

Bertola, Giuseppe, Boeri, Tito, 2003. Product Market Integration, Institutions and the Labour Markets, Mimeo. http://didattica.unibocconi.it/mypage/upload/48791_20090205_034431_48791_20090116_035648_BOERIBERTOLA14-4-03.pdf.

Bertola, Giuseppe, Blau, Francine, Kahn, Lawrence, 2001. Comparative Analysis of Labor Market Outcomes: Lessons for the US from International Long-Run Evidence, NBER Working Paper 8526.

Bertola, Giuseppe, Blau, Francine, Kahn, Lawrence, 2007. Labor market institutions and demographic employment patterns. J. Popul. Econ. 20 (4), 833–867.

Betcherman, Gordon, 2012. Labour Market Institutions. A Review of the Literature, Policy Research, Working Paper No. 6276, World Bank.

Bičáková, Alena, 2006. Market vs. Institutions: The Trade-Off Between Unemployment and Wage Inequality Revisited, EUI Working Paper ECO No. 2006/31, Florence.

Bingley, Paul, Cappellari, Lorenzo, Westergård-Nielsen, Niels, 2013. Unemployment insurance, wage dynamics and inequality over the life cycle. Econ. J. 123 (568), 341–372.

Blackburn, McKinley, Bloom, David E., 1987. Earnings and income inequality in the United States. Popul. Dev. Rev. 13 (5), 575–609.

Blackburn, McKinley, Bloom, David E., 1995. Changes in the structure of family income inequality in the United States and other industrialized nations during the 1980s. Res. Labor Econ. 14, 141–170.

Blackburn, McKinley, Bloom, David E., Freeman, Richard, 1990. An era of falling earnings and rising inequality? Brookings Rev. 9 (1), 38–43.

Blanchard, Olivier, Giavazzi, Francesco, 2003. Macroeconomic effects of regulations and deregulation in goods and labor markets. Q. J. Econ. 118 (3), 879–907.

Blanchard, Olivier, Tirole, Jean, 2008. The joint design of unemployment insurance and employment protection: a first pass. J. Eur. Econ. Assoc. 6 (1), 45–77.

Blanchard, Olivier, Wolfers, Justin, 2000. The role of shocks and institutions in the rise of European unemployment: the aggregate evidence. Econ. J. 110, C1–C33.

Blanchflower, David, Oswald, Andrew, 2013. Does High Home-Ownership Impair the Labor Market? Mimeo at andrewoswald.com.

Blau, Francine, Kahn, Lawrence, 1996. International differences in male wage inequality: institutions versus market forces. J. Polit. Econ. 104 (4), 791–837.

Blau, Francine, Kahn, Lawrence, 1999. Institutions and laws in the labor market, In: Ashenfelter, Orley, Card, David E. (Eds.), Handbook of Labor Economics. vol. 3A. pp. 1399–1461 (Chapter 25).

Blau, Francine, Kahn, Lawrence, 2002. At Home and Abroad. U.S. Labor Market Performance in International Perspective. Russell Sage, New York.

Blau, Francine, Kahn, Lawrence, 2009. Inequality and earnings distribution. In: Salverda, Wiemer, Nolan, Brian, Smeeding, Timothy (Eds.), The Oxford Handbook of Economic Inequality. Oxford University Press, Oxford, pp. 177–203 (Chapter 8).

Blinder, Alan, 2007. How Many U.S. Jobs Might Be Offshorable? Working Paper 60, Center for Economic Policy Studies, Princeton.

Blinder, Alan, Krueger, Alan, 2009. Alternative Measures of Offshorability: A Survey Approach, Working Paper 190, Center for Economic Policy Studies, Princeton.

Bluestone, Barry, Harrison, Bennett, 1982. The Deindustrialization of AMERICA. Basic Books, New York.

Bluestone, Barry, Harrison, Bennett, 1988. The Great U-Turn: Corporate Restructuring and the Polarizing of America. Basic Books, New York.

Blundell, Richard, Etheridge, Ben, 2010. Consumption, income and earnings inequality in Britain. Rev. Econ. Dyn. 13 (1), 76–102.

Boeri, Tito, 2011. Institutional reforms and dualism in European labor markets. In: Ashenfelter, Orley, Card, David (Eds.), Handbook of Labor Economics. vol. 4B. Elsevier, Amsterdam, pp. 1173–1236 (Chapter 13).

Boeri, Tito, 2012. Setting the minimum wage. Labour Econ. 19 (3), 281–290.

Boeri, Tito, Jimeno, Juan, 2005. The effects of employment protection: learning from variable enforcement. Eur. Econ. Rev. 49, 2057–2077.

Boeri, Tito, van Ours, Jan, 2008. The Economics of Imperfect Labor Markets. Oxford University Press, Oxford.

Botero, Juan C., Djankov, Simeon, La Porta, Rafael, Lopez-de-Silanes, Florencio, Shleifer, Andrei, 2004. The regulation of labor. Q. J. Econ. 119 (4), 1339–1382.

Bound, John, Johnson, George, 1989. Changes in the Structure of Wages During the 1980's: An Evaluation of Alternative Explanations, NBER Working Paper 2983.

Bound, John, Johnson, George, 1992. Changes in the structure of wages in the 1980's: an evaluation of alternative explanations. Am. Econ. Rev. 82 (3), 371–392.

Bourguignon, François, Ferreira, Francisco, Lustig, Nora (Eds.), 2004. The Microeconomics of Income Distribution Dynamics in East Asia and Latin America. World Bank and Oxford University Press, Oxford.

Bowles, Samuel, Park, Yongjin, 2005. Emulation, inequality, and work hours: was Thorsten Veblen right? Econ. J. 115, F397–F412.

Brandolini, Andrea, D'Alessio, Giovanni, 2001. Household Structure and Income Inequality, Working Paper 6/2001. Center for Household, Income, Labour and Demographic Economics CHILD.

Brandolini, Andrea, Rosolia, Alfonso, Torrini, Roberto, 2011. The Distribution of Employees' Labour Earnings in the European Union: Data, Concepts and First Results, Working Paper 2011-198, ECINEQ.

Brewer, Mike, Wren-Lewis, Liam, 2012. Why did Britain's households get richter? Decomposing UK Household Income Growth Between 1968 and 2008–09, Working Paper 2012–08, ISER, University of Essex.

Brewer, Mike, Duncan, Alan, Shephard, Andrew, Suarez, María José, 2006. Did working families' tax credit work? the impact of in-work support on labour supply in Great Britain. Labour Econ. 13, 699–720.

Brewer, Mike, Muriel, Alastair, Wren-Lewis, Liam, 2009. Accounting for Changes in Inequality Since 1968: Decomposition Analyses for Great Britain, Mimeo.

Brown, Charles, 1999. Minimum wages, employment, and the distribution of income. In: Ashenfelter, Orley, Card, David (Eds.), Handbook of Labor Economics. vol. 3. Elsevier, Amsterdam, pp. 2101–2163.

Brügemann, Björn, 2012. Does employment protection create its own political support? J. Eur. Econ. Assoc. 10 (2), 369–416.

Bryson, Alex, Freeman, Richard, Lucifora, Claudio, Pellizzari, Michele, Perotin, Virginie, 2012. Paying for Performance: Incentive Pay Schemes and Employees' Financial Participation, CEP Discussion Paper 1112, LSE.

Brzozowski, Michael, Gervais, Martin, Klein, Paul, Suzuki, Michio, 2010. Consumption, income, and wealth inequality in Canada. Rev. Econ. Dyn. 13 (1), 52–75.

Budría, Santiago, Pereira, Pedro T., 2005. Educational Qualifications and Wage Inequality: Evidence for Europe, IZA Discussion Paper 1763. Bonn.

Burtless, Gary, 1999. Effects of growing wage disparities and changing family composition on the U.S. income distribution. Eur. Econ. Rev. 43, 853–865.

Burtless, Gary, 2009. Demographic transformation and economic inequality. In: Salverda, Weimer, Nolan, Brian, Smeeding, Timothy (Eds.), The Oxford Handbook of Economic Inequality. Oxford University Press, Oxford, pp. 435–454 (Chapter 18).

Butcher, Tim, 2011. Still evidence-based? The role of policy evaluation in recession and beyond: the case of the national minimum wage. Natl. Inst. Econ. Rev. 219, R26–R40.

Butcher, Tim, Dickens, Richard, Manning, Alan, 2012. Minimum Wages and Wage Inequality: Some Theory and an Application to the UK, CEP Discussion Paper 1177, LSE.

Cancian, Maria, Schoeni, Robert, 1998. Wives' earnings and the level and distribution of married couples earnings in developed countries. J. Income Distribut. 8 (1), 45–61.

Card, David, DiNardo, John, 2002. Skill-biased technological change and rising wage inequality: some problems and puzzles. J. Labor Econ. 20 (4), 733–783.

Card, David, Krueger, Alan, 1995a. Myth and Measurement. The New economics of the Minimum Wage. Princeton University Press, Princeton.

Card, David, Krueger, Alan, 1995b. Time-series minimum-wage studies: a meta-analysis. Am. Econ. Rev. 85, 238–243.

Card, David, Lemieux, Thomas, Riddell, Craig, 2004. Unions and wage inequality. J. Labor Res. 25 (4), 519–562.

Cardoso, Ana, 2010. Do firms compress the wage distribution? In: Marsden, David, Rycx, François (Eds.), Wage Structures, Employment Adjustments and Globalization: Evidence from Linked and Firm-level Panel Data. Palgrave Macmillan, Basingstoke, pp. 202–218.
Charles, Kerwin Kofi, Guryan, Jonathan, 2007. Prejudice and the Economics of Discrimination, NBER Working Paper 13661.
Charles, Kerwin Kofi, Guryan, Jonathan, 2008. Prejudice and the economics of discrimination. J. Polit. Econ. 116 (5), 773–809.
Checchi, Daniele, Lucifora, Claudio, 2002. Unions and labour market institutions in Europe. Econ. Policy 17 (2), 362–401.
Checchi, Daniele, Garcia Peñalosa, Cecilia, 2008. Labour market institutions and income inequality. Econ. Policy 56, 600–651.
Checchi, Daniele, Garcia Peñalosa, Cecilia, 2010. Labour shares and the personal distribution of income in the OECD. Economica 77 (307), 413–450.
Cheon, Byung You, Chang, Jiyeun Hwang, Seong Shin, Gyu, Wook Kang, Jin, Wook Lee, Shin, Hee Kim, Byung, Joo, Hyun, 2013. Growing Inequality and Its Impacts in Korea, Country Report to the GINI project, http://gini-research.org/CR-Korea.
Chetty, Raj, 2008. Moral Hazard vs. Liquidity and Optimal Unemployment Insurance, NBER Working Paper 13967.
Christopoulou, Rebekka, Jimeno, Juan F., Lamo, Ana, 2010. Changes in the Wage Structure in EU Countries, Documentos de Trabajo 1017. Banco de España.
Coelli, Michael, Fahrer, Jerome, Lindsay, Holly, 1994. Wage Dispersion and Labour Market Institutions: A Cross Country Study, Research Discussion Paper 9404. Reserve Bank of Australia.
Congressional Budget Office (CBO), 2012. The Distribution of Household Income and Average Federal Tax Rates, 2008 and 2009. Supplemental data, http://www.cbo.gov/publication/43373.
Conyon, Martin, Fernandes, Nuno, Ferreira, Miguel, Matos, Pedro, Murphy, Kevin, 2011. The Executive Compensation Controversy: A Transatlantic Analysis, Fondazione Rodolfo De Benedetti.
Corluy, Vincent, Vandenbroucke, Frank, 2013. Individual Employment, Household Employment and Risk of Poverty in the EU. A Decomposition Analysis, Eurostat.
Corsini, Lorenzo, 2008. Institutions, Technological Change and the Wage Differentials Between Skilled and Unskilled Workers: Theory and Evidence from Europe, IRISS Working Paper 2008-02.
Cowell, Frank, 2000. Measurement of inequality, In: Atkinson, A.B., Bourguignon, F. (Eds.), Handbook of Income Distribution. vol. 1. pp. 87–166, Elsevier, Amsterdam (Chapter 2).
Cowell, Frank, Fiorio, Carlo, 2011. Inequality decompositions—a reconciliation. J. Econ. Inequality 9, 509–528.
Dahl, Christian, le Maire, Daniel, Munch, Jakob, 2011. Wage Dispersion and Decentralization of Wage Bargaining, IZA Discussion Paper 6176.
Daly, Mary, Valetta, Robert, 2006. Inequality and poverty in United States: the effects of rising dispersion of men's earnings and changing family behaviour. Economica 73 (289), 75–98.
De Graaf-Zijl, Marloes, Nolan, Brian, 2011. Household joblessness and its impact on poverty and deprivation in Europe. J. Eur. Soc. Policy 21, 413–431.
De Linde Leonard, M., Stanley, T.D., Doucouliagos, Hristos, 2014. Does the UK minimum wage reduce employment? A meta-regression analysis. Br. J. Indus. Relat. 52 (3), 499–520.
Deininger, Klaus, Squire, Lyn, 1996. A new data set measuring income inequality. World Bank Econ. Rev. 10 (3), 565–591.
Del Boca, Daniela, Pasqua, Sylvia, 2003. Employment patterns of husbands and wives and family income distribution in Italy (1977–98). Rev. Income Wealth 49 (2), 221–245.
DiNardo, John, Lemieux, Thomas, 1997. Diverging male wage inequality in the United States and Canada, 1981–1988: do institutions explain the difference? Indust. Labor Relat. Rev. 50 (4), 629–651.
DiNardo, John, Pischke, Jorn-Steffen, 1997. The returns to computer use revisited: have pencils changed the wage structure too? Q. J. Econ. 112 (1), 291–303.
DiNardo, John, Fortin, Nicole, Lemieux, Thomas, 1996. Labor market institutions and the distribution of wages, 1973–1992: a semiparametric approach. Econometrica 64 (5), 1001–1044.

Dingeldey, Irene, 2001. European tax systems and their impact on family employment patterns. J. Soc. Policy 30, 653–672.
DiPrete, Thomas, 2007. What has sociology to contribute to the study of inequality trends? An historical and comparative perspective. Am. Behav. Sci. 50 (5), 603–618.
Dolado, Juan, Kramarz, Francis, Machin, Stephen, Manning, Alan, Margolis, David, Teulings, Coen, 1996. The economic impact of minimum wage in Europe. Econ. Policy 23, 317–372.
Dolton, Peter, Bondibene, Chiara Rosazza, 2011. An evaluation of the international experience of minimum wages in an economic downturn, Research Report for the Low Pay Commission, University of London, Royal Holloway College.
Domeij, David, Floden, Martin, 2010. Inequality trends in Sweden 1978–2004. Rev. Econ. Dyn. 13 (1), 179–208.
Dooley, Martin, Gottschalk, Peter, 1984. Earnings inequality among males in the United States: trends and the effect of labor force. J. Polit. Econ. 92 (1), 59–89.
Dooley, Martin, Gottschalk, Peter, 1985. The increasing proportion of men with low earnings in the United States. Demography 22 (1), 25–34.
Doucouliagos, Hristos, Stanley, T.D., 2009. Publication selection bias in minimum-wage research? A meta-regression analysis. Br. J. Ind. Relat. 47 (22), 406–428.
Dube, Arindrajit, 2013. Minimum Wages and the Distribution of Family Incomes. https://dl.dropboxusercontent.com/u/15038936/Dube_MinimumWagesFamilyIncomes.pdf (December 30, 2013).
Dube, Arindrajit, Lester, William, Reich, Michael, 2010. Minimum wage effects across state borders: estimates using contiguous counties. Rev. Econ. Stat. 92 (4), 945–964.
Dube, Arindrajit, Lester, William, Reich, Michael, 2012. Minimum Wage Shocks, Employment Flows and Labor Market Frictions. Institute for Research on Labor and Employment, Berkeley, CA.http://escholarship.org/uc/item/76p927ks.
Dunne, Timothy, Foster, Lucia, Haltiwanger, John, Troske, Kenneth R., 2004. Wage and productivity dispersion in United States manufacturing: the role of computer investment. J. Labor Econ. 22 (2), 397–429.
Dupuy, Arnaud, Fernández-Kranz, Daniel, 2011. International differences in the family gap in pay: the role of labour market institutions. Appl. Econ. 43, 413–438.
Dustmann, Christian, Ludsteck, Johannes, Schönberg, Uta, 2009. Revisiting the German wage structure. Q. J. Econ. 124 (2), 843–881.
Eichhorst, Werner, Feil, Michael, Braun, Christoph, 2008. What Have We Learned? Assessing Labor Market Institutions and Indicators, IZA Discussion Paper 3470. Bonn.
Eissa, Nada, Hoynes, Hilary, 2004. Taxes and the labor market participation of married couples: the earned income tax credit. J. Public Econ. 88, 1931–1958.
Eissa, Nada, Hoynes, Hilary, 2006. Behavioral responses to taxes: lessons from the EITC and labor supply. In: Poterba, J.M. (Ed.), Tax Policy and the Economy. MIT Press, Cambridge MA, pp. 73–110 (Chapter 3).
Farber, Henry, Valletta, Robert, 2013. Do extended Unemployment Benefits Lengthen Unemployment Spells? Evidence From Recent Cycles in the US Labour Market, Working Paper 573, Princeton University, Industrial Relations section.
Fields, Gary, 2003. Accounting for income inequality and its change: a new method, with application to the distribution of earnings in the United States. Res. Labor Econ. 22, 1–38.
Fiori, Giuseppe, Nicoletti, Giuseppe, Scarpetta, Stefano, Schiantarelli, Fabio, 2012. Employment effects of product and labour market reforms: are there synergies? Econ. J. 122, F79–F104.
Firpo, Sergio, Fortin, Nicole, Lemieux, Thomas, 2009. Unconditional quantile regressions. Econometrica 77 (3), 953–973.
Firpo, Sergio, Fortin, Nicole M., Lemieux, Thomas, 2011. Occupational Tasks and Changes in the Wage Structure, IZA Discussion Paper 5542. Bonn.
Flaig, Gebhard, Rottmann, Horst, 2011. Labour Market Institutions and Unemployment. An International Comparison, CESifo Working Paper 3558.
Förster, Michael, 1994. Measurement of Low Incomes and Poverty in a Perspective of International Comparisons, OECD Labour Market and Social Policy Occasional Papers 14.

Fortin, Nicole, Lemieux, Thomas, 1997. Institutional changes and rising wage inequality: is there a linkage? J. Econ. Perspect. 1 (2), 75–96.

Fortin, Nicole, Lemieux, Thomas, 2014. Changes in wage inequality in Canada: an interprovincial perspective, First draft, http://faculty.arts.ubc.ca/nfortin/FortinLemieux_Inequality_%20provinces.pdf.

Fortin, Nicole, Lemieux, Thomas, Firpo, Sergio, 2011. Decomposition methods in economics. In: Ashenfelter, Orley, Card, David (Eds.), Handbook of Labor Economics. vol. IIIA. Elsevier, Amsterdam, pp. 1463–1555. (Chapter 1; also NBER Working Paper 16045, 2010).

Fortin, Nicole M., Schirle, Tammy D., 2006. Gender dimensions of changes in earnings inequality in Canada. In: Green, David A., Kesselman, Jonathan R. (Eds.), Dimensions of Inequality in Canada. Vancouver, UBC Press, pp. 307–346.

Freeman, Richard, 1980. Unionism and the dispersion of wages. Ind. Labor Relat. Rev. 34 (2), 3–23.

Freeman, Richard, 1988. Labour market institutions and economic performance. Econ. Policy 3 (6), 63–88.

Freeman, Richard, 1991. How Much Has De-Unionisation Contributed to the Rise in Male Earnings Inequality? NBER Working Paper 3826 (In: Danziger and Gottschalk (Eds.), Uneven Tides, 1992).

Freeman, Richard (Ed.), 1994. Working Under Different Rules. Russell Sage, New York.

Freeman, Richard, 2000. Single Peaked vs. Diversified Capitalism: The Relation Between Economic Institutions and Outcomes, NBER Working Paper 7556.

Freeman, Richard B., 2001. Single Peaked Vs. Diversified Capitalism: The Relation Between Economic Institutions and Outcomes, Working Paper No. 7556, NBER, Boston.

Freeman, Richard, 2005. Labour Market Institutions without Blinders: The Debate over Flexibility and Labour Market Performance, NBER Working Paper 11286.

Freeman, Richard, 2006. The great doubling: the challenge of the new global labor market. http://emlab.berkeley.edu/users/webfac/eichengreen/e183_sp07/great_doub.pdf.

Freeman, Richard, 2007. Labor market institutions around the world, NBER Working Paper 13242.

Freeman, Richard, 2009. Labor Regulations, Unions, and Social Protection in Developing Countries: Market Distortions or Efficient Institutions? NBER Working Paper 14789.

Freeman, Richard, Katz, Lawrence (Eds.), 1995. Differences and Changes in Wage Structures. University of Chicago Press, Chicago.

Freeman, Richard, Medoff, James, 1984. What Do Unions Do? Basic Books, New York.

Freeman, Richard, Schettkat, Ronald, 2001. Skill compression, wage differentials and employment: Germany vs. the US. Oxf. Econ. Pap. 53 (3), 582–603.

Friedman, Milton, 1968. The role of monetary policy. Am. Econ. Rev. 58 (1), 1–17.

Fuchs-Schündeln, Nicola, Krueger, Dirk, Sommer, Mathias, 2010. Inequality trends for Germany in the last two decades: a tale of two countries. Rev. Econ. Dyn. 13 (1), 103–132.

Garnero, Andrea, Kampelmann, Stephan, Rycx, François, 2013. Minimum Wage Systems and Earnings Inequalities: Does Institutional Diversity Matter? DULBEA Working Paper 13-06.

Gautié, Jérôme, Schmitt, John (Eds.), 2010. Low-Wage Work in Wealthy Countries. Russell Sage, New York.

Gindling, Thomas, Terrell, Katherine, 2009. Minimum wages, wages and employment in various sectors in Honduras. Labour Econ. 16 (3), 291–303.

Giuliano, Laura, 2013. Minimum wage effects on employment, substitution, and the teenage labor supply: evidence from personnel data. J. Labor Econ. 31 (1), 155–194.

Glyn, Andrew, 2009. Functional distribution and inequality. In: Salverda, Weimer, Nolan, Brian, Smeeding, Timothy M. (Eds.), Oxford Handbook of Economic Inequality. Oxford University Press, Oxford, pp. 101–126 (Chapter 5).

Glyn, Andrew, Baker, Dean, Howell, David, Schmitt, John, 2003. Labor Market Institutions and Unemployment: A Critical Review of the Cross-Country Evidence, Discussion Paper 168. Department of Economics, University of Oxford, Oxford.

Golden, Miriam, Wallerstein, M., 2006. Domestic and International Causes for the Rise of Pay Inequality: Post -Industrialism, Globalization and Labor Market Institutions. Mimeo.

Golden, Miriam, Wallerstein, Michael, 2011. Domestic and international causes for the rise of pay inequality in OECD nations between 1980 and 2000. In: Comparing European Workers. Part A: Experiences and Inequalities. In: Research in the Sociology of Work, vol. 22. pp. 209–249.

Goos, Maarten, Manning, Alan, 2003. Lousy and Lovely Jobs: the Rising Polarization of Work in Britain. London, LSE, CEP Working Paper 604.
Goos, Maarten, Manning, Alan, 2007. Lousy and lovely jobs: the rising polarization of work in Britain, Rev. Econ. Stat. 89 (1), 118–133.
Goos, Maarten, Manning, Alan, Salomons, Anna, 2009. Job polarization in Europe. Am. Econ. Rev. 99 (2), 58–63.
Goos, Maarten, Manning, Alan, Salomons, Anna, 2010. Explaining Job Polarization in Europe: The Roles of Technology, Globalization and Institutions, CEP Discussion Paper 1026, LSE.
Goos, Maarten, Manning, Alan, Salomons, Anna, 2011. Explaining Job Polarization: The Roles of Technology, Offshoring and Institutions. KU Leuven, Center for Economic Studies, Discussion Paper 11.34.
Goos, Maarten, Manning, Alan, Salomons, Anna, 2014. Explaining job polarization: routine-biased technological change and offshoring. Am. Econ. Rev. 104 (8), 2509–2526.
Gottschalk, Peter, Danziger, Sheldon, 2005. Inequality of wage rates, earnings and family income in the United States 1975–2002. Rev. Income Wealth 51 (2), 231–254.
Gottschalk, Peter, Joyce, Mary, 1998. Cross-national differences in the rise in earnings inequality: market and institutional factors. Rev. Econ. Stat. 80 (4), 489–502.
Gottschalk, Peter, Smeeding, Timothy, 1997. Cross-national comparisons of earnings and income inequality. J. Econ. Lit. 35, 633–687.
Gregg, Paul, Wadsworth, Jonathan, 1996. More work in fewer households? In: Hills, J. (Ed.), New Inequalities. Cambridge University Press, London, pp. 181–207.
Gregg, Paul, Wadsworth, Jonathan, 1998. It Takes Two: Employment Polarisation in the OECD, Discussion Paper 304, Centre for Economic Performance, London.
Gregg, Paul, Wadsworth, Jonathan, 2008. Two sides to every story: measuring polarization and inequality in the distribution of work. J. R. Stat. Soc. Ser. A 171 (4), 857–875.
Gregg, Paul, Scutella, Rosanna, Wadsworth, Jonathan, 2010. Reconciling workless measures at the individual and household level. Theory and evidence from the United States, Britain, Germany, Spain and Australia. J. Popul. Econ. 23, 139–167.
Hall, Robert, Krueger, Alan, 2010. Evidence on the Determinants of the Choice Between Wage Posting and Wage Bargaining, NBER Working Paper 16033.
Hall, Peter, Soskice, David (Eds.), 2001. Varieties of Capitalism: The Institutional Foundations of Comparative Advantage. Oxford University Press, Oxford.
Harrison, Bennet, Tilly, Chris, Bluestone, Barry, 1986. Wage inequality takes a great U-turn. Challenge 29 (1), 26–32, Symposium of the Joint Economic Committee of Congress on the 40th Anniversary of the Employment Act of 1946 (Part 1).
Heathcote, Jonathan, Perri, Fabrizio, Violante, Giovanni, 2010. Unequal we stand: an empirical analysis of economic inequality. Rev. Econ. Dyn. 13, 15–51.
Heckman, James, Sedlacek, Guilherme, 1985. Heterogeneity, aggregation, and market wage functions: an empirical model of self-selection in the labor market. J. Polit. Econ. 93 (9), 1077–1125.
Henle, Peter, 1972. Distribution of earned income. Monthly Labor Rev. 95 (12), 16–27.
Henle, Peter, Ryscavage, Paul, 1980. The distribution of earned income among men and women, 1958–77. Monthly Labour Rev. 103 (4), 3–10.
Heshmati, Almas, 2004. A Review of Decomposition of Income Inequality, Discussion Paper 1221, IZA Bonn.
Hills, John, Brewer, Mike, Jenkins, Stephen, Lister, Ruth, Lupton, Ruth, Machin, Stephen, Mills, Colin, Modood, Tariq, Rees, Teresa, Riddell, Sheila, 2010. An Anatomy of Economic Inequality in the UK: Report of the National Equality Panel. http://sticerd.lse.ac.uk/case/_new/publications/NEP.asp.
Holzmann, Robert, Pouget, Yann, Vodopivec, Milan, Weber, Michael, 2011. Severance Pay Programs around the World: History, Rationale, Status, and Reforms, IZA Discussion Paper 5731.
Hyslop, Dean, 2001. Rising US earnings inequality and family labor supply: the covariance structure of intrafamily earnings. Am. Econ. Rev. 91, 755–777.
ILO, 2008. Global Wage Report 2008/09. Minimum Wages and Collective Bargaining: Towards Policy Coherence, Geneva.
ILO, 2010. Global Wage Report 2010/11. Wage Policies in Times of Crisis, Geneva.

Japelli, Tullio, Pistaferri, Luigi, 2010. Does consumption inequality track income inequality in Italy? Rev. Econ. Dyn. 13 (1), 133–153.

Jenkins, Stephen, 1995. Accounting for inequality trends: decomposition analyses for the UK, 1971–86. Economica 62 (245), 29–63.

Jenkins, Stephen, van Kerm, Philippe, 2009. The measurement of economic inequality. In: Salverda, Weimer, Nolan, Brian, Smeeding, Timothy (Eds.), The Oxford Handbook of Economic Inequality. Oxford University Press, Oxford, pp. 40–67 (Chapter 3).

Johnson, David, Wilkins, Roger, 2003. The Effects of Changes in Family Composition and Employment Patterns on the Distribution of Income in Australia: 1982 to 1997–1998, Working Paper 19/03, Melbourne Institute of Applied Economic and Social Research.

Juhn, Chinhui, Murphy, Kevin, Pierce, Brooks, 1993. Wage inequality and the rise in returns to skill. J. Polit. Econ. 101, 410–442.

Kahn, James, 2005. Labor Supply and the Changing Household. Paper presented to Society for Economic Dynamics.

Kampelmann, Stephan, Garnero, Andrea, Rycx, François, 2013. Minimum Wages in Europe: Does the Diversity of Systems Lead to a Diversity of Outcomes? ETUI Report 128, Brussels.

Karabarbounis, Loukas, Neiman, Brent, 2013. The Global Decline of the Labor Share, NBER Working Paper 19136.

Karoly, Lynn, 1988. A Study of the Distribution of Individual Earnings in the United States from 1967 to1986. PhD Thesis, Yale University.

Karoly, Lynn, 1992. Changes in the distribution of individual earnings in the United States: 1967–1986. Rev. Econ. Stat. 74 (1), 107–115.

Karoly, Lynn, 1993. The trend in inequality among families, individuals and workers in the United States: a twenty-five year perspective, In: Danziger, S., Gottschalk, P. (Eds.), Uneven Tides, Rising Inequality in America. Russell Sage, New York, pp. 19–97 (Chapter 2).

Karoly, Lynn, Burtless, Gary, 1995. Demographic change, rising earnings inequality, and the distribution of personal well-being, 1959–1989. Demography 32 (3), 379–405.

Katz, Lawrence, Autor, David, 1999. Changes in the Wage Structure and Earnings Inequality. In: Ashenfelter, Orley, Card, David (Eds.), Handbook of Labor Economics. vol. IIIA. Elsevier, Amsterdam, pp. 1463–1555.

Katz, Lawrence, Murphy, Kevin, 1992. Changes in relative wages: supply and demand factors. Q. J. Econ. CVII, 35–78.

Kenworthy, Lane, 2001. Wage-setting measures: a survey and assessment. World Polit. 54, 57–98.

Kenworthy, Lane, 2008. Sources of Equality and Inequality: Wages, Jobs, Households, and redistribution, Working Paper 471, Luxembourg Income Study. Chapter 3 of Kenworthy, Jobs with Equality, Oxford University Press, 2008.

Kenworthy, Lane, Pontusson, Jonas G, 2005. Rising inequality and the politics of redistribution in affluent countries. Perspect. Polit. 3 (3), 449–471.

Kierzenkowski, Rafal, Koske, Isabel, 2012. Less Income Inequality and More Growth—Are they Compatible? Part 8. The Drivers of Labour Income Inequality—A Literature Review, OECD Economics Department Working Paper 931, OECD Publishing.

Kimhi, Ayal, 2011. On the interpretation (and misinterpretation) of inequality decompositions by income sources. World Dev. 39 (10), 1888–1890.

Kluve, Jochen, 2010. The effectiveness of European active labor market programs. Labour Econ. 17 (6), 904–918.

Koeniger, Winfried, Leonardi, Marco, Nunziata, Luca, 2007. Labor market institutions and wage inequality. Ind. Labor Relat. Rev. 60 (3), 340–356.

Krueger, Dirk, Perri, Fabrizio, 2006. Does income inequality lead to consumption inequality? Evidence and theory. Rev. Econ. Stud. 73 (1), 163–193.

Kugler, Adriana, Pica, Giovanni, 2008. Effects of employment protection on worker and job flows: evidence from the 1990 Italian reform. Labour Econ. 15 (1), 78–95.

Lalive, Rafael, Schlosser, Analía, Steinhauer, Andreas, Zweimüller, Josef, 2011. Parental Leave and Mothers' Careers: The Relative Importance of Job Protection and Cash Benefits, IZA Discussion Paper 5792, Bonn.

Lane, Julia, 2009. Inequality and the labour market – employers. In: Salverda, Wiemer, Nolan, Brian, Smeeding, Timothy (Eds.), The Oxford Handbook on Economic Inequality. Oxford University Press, Oxford, pp. 204–229 (Chapter 9).
Larrimore, Jeff, 2013. Accounting for United States household income inequality trends: the changing importance of household structure and male and female labour earnings inequality. Rev. Income Wealth 101 (3), 173–177.
Layard, Richard, Nickell, Stephen, Jackman, Richard, 1991. Unemployment. Macroeconomic Performance and the Labour Market. Oxford University Press, Oxford.
Lee, David, 1999. Wage inequality in the United States during the 1980s: rising dispersion or falling minimum wage? Q. J. Econ. 114 (3), 977–1023.
Lehrer, Evelyn, 2000. The impact of women's employment on the distribution of earnings among married-couple households: a comparison between 1973 and 1992–1994. Q. Rev. Econ. Finance 40, 295–301.
Leigh, Andrew, 2009. Top incomes, In: Salverda, Wiemer, Nolan, Brian, Smeeding, Timothy (Eds.), The Oxford Handbook of Economic Inequality. Oxford University Press, Oxford, pp. 150–174 (Chapter 7).
Lemieux, Thomas, 2006a. Postsecondary education and increasing wage inequality. Am. Econ. Rev. 96 (2), 195–199.
Lemieux, Thomas, 2006b. Postsecondary Education and Increasing Wage Inequality, NBER Working Paper 12077.
Lemieux, Thomas, 2006c. Increasing residual wage inequality: composition effects, noisy data, or rising demand for skill? Am. Econ. Rev. 96 (3), 461–498.
Lemieux, Thomas, 2008. The changing nature of wage inequality. J. Popul. Econ. 21 (1), 21–48.
Lemieux, Thomas, 2010. What Do We Really Know about Changes in Wage Inequality?. In: Abraham, Katharine, Spletzer, James, Harper, Michael (Eds.), Labor in the New Economy. NBER Books, University of Chicago Press, Boston/Chicago, pp. 17–59 (Chapter 1).
Lemieux, Thomas, 2011. Wage inequality: a comparative perspective. Aust. Bull. Labour 37 (1), 2–32.
Lemieux, Thomas, Macleod, Bentley, Parent, Daniel, 2009. Performance pay and wage inequality. Q. J. Econ. 124 (1), 1–49.
Lemos, Sara, 2009. Minimum wage effects in a developing country. Labour Econ. 16, 224–237.
Leonardi, Marco, Pica, Giovanni, 2013. Who pays for it? The heterogeneous wage effects of employment protection legislation. Econ. J. 123 (573), 1236–1278.
Lerman, Robert, Yitzhaki, Shlomo, 1984. A note on the calculation and interpretation of the Gini index. Econ. Lett. 15 (3–4), 363–368.
Leuven, Edwin, Oosterbeek, Hessel, van Ophem, Hans, 1997. International Comparisons of Male Wage Inequality; Are the Findings Robust? Tinbergen Institute, Working Paper 97-59, University of Amsterdam.
Leuven, Edwin, Oosterbeek, Hessel, van Ophem, Hans, 2004. Explaining international differences in male skill wage differentials by differences in demand and supply of skill. Econ. J. 114, 466–486.
Levy, Frank, Murnane, Richard, 1992. U.S. earnings levels and earnings inequality: a review of recent trends and proposed explanations. J. Econ. Lit. 30, 1333–1381.
Levy, Frank, Murnane, Richard, 2005. How Computerized Work and Globalization Shape Human Skill Demands, MIT IPC Working Paper IPC-05-006.
Levy, Frank, Temin, Peter, 2007. Inequality and Institutions in 20th Century America, NBER Working Paper 13106.
Liebman, Jeffrey, 1998. The impact of the earned income tax credit on incentives and income distribution. In: Poterba, James (Ed.), Tax Policy and the Economy. vol. 12.
Liu, Yujia, Grusky, David, 2013. The payoff to skill in the third industrial revolution. Am. J. Sociol. 118 (5), 1330–1374.
Lu, Yuqian, Morissette, René, Schirle, Tammy, 2011. The growth of family earnings inequality in Canada 1980–2005. Rev. Income Wealth 57 (1), 23–39.
Lucifora, Claudio, Salverda, Wiemer, 2009. Low pay, In: Salverda, Wiemer, Nolan, Brian, Smeeding, Timothy (Eds.), The Oxford Handbook of Economic Inequality. Oxford University Press, Oxford, pp. 256–283 (Chapter 11).
Lydall, Harold, 1968. The Structure of Earnings. Clarendon Press, Alderley, Gloucestershire.

Machin, Stephen, 1997. The decline of labour market institutions rise in wage inequality in Britain. Eur. Econ. Rev. 41 (3–5), 647–657.

Machin, Stephen, 2008. An appraisal of economic research on changes in wage inequality. Labour 22 (Special issue), 7–26.

Mandel, Hadas, Semyonov, Moshe, 2005. Family policies, wage structures, and gender gaps: sources of earnings inequality in 20 countries. Am. Sociol. Rev. 70, 949–967.

Mandelbrot, Benoit, 1962. Paretian distributions and income maximization. Q. J. Econ. 76 (1), 57–85.

Manning, Alan, 2011. Imperfect competition in the labour market. In: Ashenfelter, Orley, Card, David (Eds.), Handbook of Labor Economics. vol. 4B. Elsevier, Amsterdam, pp. 973–1041.

Manzo IV, Frank, Bruno, Robert, 2014. Which Labor Market Institutions Reduce Income Inequality? Labor Unions, Prevailing Wage Laws, and Right-to-Work Laws in the Construction Industry, Illinois Economic Policy Institute (ILEPI). Research Report.

Martin, John, Scarpetta, Stefano, 2011. Setting It Right: Employment Protection, Labour Reallocation and Productivity, IZA Policy Paper 27, 27.

Martins, Pedro, 2009. Dismissals for cause: the difference that just eight paragraphs can make. J. Labor Econ. 27 (2), 257–279.

Matano, Alessia, Naticchioni, Paolo, 2011. Is there rent sharing in Italy? evidence from employer-employee data. Eur. J. Comp. Econ. 8 (2), 265–279.

Messina, Julian, Vallanti, Giovanna, 2007. Job flow dynamics and firing restrictions: evidence from Europe. Econ. J. 117, F279–F301.

Metcalf, David, 2004. The impact of the national minimum wage on the pay distribution, employment and training. Econ. J. 114, C84–C86.

Mishel, Lawrence, Shierholdz, Heidi, Schmitt, John, 2013. Don't Blame the Robots. Assessing the Job Polarization Explanation of Growing Wage Inequality, Working Paper, EPI, Washington (19 November).

Mookherjee, Dilip, Shorrocks, Anthony, 1982. A decomposition analysis of the trend in UK income inequality. Econ. J. 92 (368), 886–902.

Mortensen, Dale, 2005. Wage Dispersion: Why Are Similar Workers Paid Differently? MIT Press, Cambridge MA.

Mühlau, Peter, Horgan, Justine, 2001. Labour Market Status and the Wage Position of the Low Skilled: The Role of Institutions and of Demand and Supply—Evidence from the International Adult Literacy Survey, Working Paper 5, European Low-wage employment Research network. http://www.uva-aias.net/uploaded_files/regular/05MuehlauHorgan.pdf.

Neal, Derek, Rosen, Sherwin, 2000. Theories of the distribution of earnings, In: Atkinson, A., Bourguignon, F. (Eds.), Handbook of Income Distribution, pp. 379–428, Elsevier, Amsterdam.

Neumark, David, Wascher, William, 2008. Minimum Wages. MIT Press, Cambridge MA.

Neumark, David, Schweitzer, Mark, Wascher, William, 2004. Minimum wage effects throughout the wage distribution. J. Hum. Resour. 39 (2), 425–445.

Neumark, David, Salas, Ian, Wascher, William, 2013. Revisiting the Minimum Wage-Employment Debate: Throwing Out the Baby with the Bathwater? NBER Working Paper 18681.

Nickell, Stephen, 1997. Unemployment and labor market rigidities: Europe versus North America. J. Econ. Perspect. 11, 55–74.

Nickell, Stephen, Layard, Richard, 1999. Labour market institutions and economic performance, In: Ashenfelter, Orley, Card, David (Eds.), pp. 3029–3084 (Chapter 46).

Nickell, Stephen, Nunziata, Luca, Ochel, Wolfgang, 2005. Unemployment in the OECD Since the 1960s. What do we know? Econ. J. 115 (500), 1–27.

Nunziata, Luca, 2005. Institutions and wage determination: a multi-country approach. Oxf. Bull. Econ. Stat. 67 (4), 435–466.

Ochel, Wolfgang. 2005. Concepts and Measurement of Labour Market Institutions. CESifo DICE Report 4/2005, pp. 40–55.

OECD, 2004. Employment Outlook. OECD Publishing, Paris.

OECD, 2009. What Are Equivalence Scales?. OECD Publishing, Paris.

OECD, 2011. Divided We Stand: Why Inequality Keeps Rising. OECD Publishing, Paris.

OECD, 2012. Employment Outlook. OECD Publishing, Paris.

OECD, 2013. OECD Skills Outlook 2013. First Results from the Survey of Adult Skills. OECD Publishing, Paris.
Oliver, Rebecca, 2008. Diverging developments in wage inequality which institutions matter? Comp. Polit. Stud. 41 (12), 1551–1582.
Olivetti, Claudia, Petrongolo, Barbara, 2008. Unequal pay or unequal employment? A cross-country analysis of gender gaps. J. Labor Econ. 26 (4), 621–654.
Oswald, Andrew, 1996. A Conjecture on the Explanation for High Unemployment in the Industrialised Nations, Warwick Economic Research Papers 475.
Paul, Satya, 2004. Income sources effects on inequality. J. Dev. Econ. 73 (1), 435–451.
Peichl, Andreas, Pestel, Nico, Schneider, Hilmar, 2010. Does Size Matter? The Impact of Changes in Household Structure on Income Distribution in Germany, CESifo Working Paper 3219.
Philippon, T., 2001. The impact of differential payroll tax subsidies on minimum wage employment. J. Public Econ. 82, 115–146.
Pierce, Brooks, 2001. Compensation inequality. Q. J. Econ. 116 (4), 1493–1525.
Pierce, Brooks, 2010. Recent trends in compensation inequality. In: Abraham, Katharine, Spletzer, James, Harper, Michael (Eds.), Labor in the New Economy. NBER Books, University of Chicago Press, Boston/Chicago, pp. 63–98 (Chapter 2).
Pijoan-Mas, Josep, Sánchez-Marcos, Virginia, 2010. Spain is different: falling trends of inequality. Rev. Econ. Dyn. 13 (1), 154–178.
Piketty, Thomas, 2003. Income inequality in France, 1901–1998. J. Polit. Econ. 111, 1004–1042.
Piketty, Thomas, Saez, Emmanuel, 2003. Income inequality in the United States, 1913–1998. Q. J. Econ. 118, 1–39.
Piketty, Thomas, Saez, Emmanuel, 2006. The evolution of top incomes: a historical and international perspective. Am. Econ. Rev. 96 (2), 200–205.
Piketty, Thomas, Saez, Emmanuel, 2012. Optimal Labor Income Taxation, NBER Working Paper18521.
Piketty, Thomas, Saez, Emmanuel, 2013. Optimal labor income taxation, vol. 5. In: Auerbach, Alan, Chetty, Raj, Feldstein, Martin, Saez, Emmanuel (Eds.), Handbook of Public Economics. Elsevier, Amsterdam (Chapter 7).
Piketty, Thomas, Saez, Emmanuel, Stantcheva, Stefanie, 2011. Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities, NBER Working Paper17616.
Plotnick, Robert, 1982. Trends in male earnings inequality. South. Econ. J. 48 (3), 724–732.
Reed, Deborah, Cancian, Maria, 2001. Sources of inequality: measuring the contributions of income sources to rising family income inequality. Rev. Income Wealth 47 (3), 321–333.
Rogerson, Richard, Shimer, Robert, 2011. Search in macroeconomic models of the labor market. In: Ashenfelter, Orley, Card, David (Eds.), Handbook of Labor Economics. vol. 4A. Elsevier, Amsterdam, pp. 619–700.
Rogerson, Richard, Shimer, Robert, Wright, Randall, 2005. Search-theoretic models of the labor market: a survey. J. Econ. Lit. 43 (4), 959–988.
Rosen, Sherwin, 1986. The theory of equalizing differences. In: Ashenfelter, Orley, Layard, Richard (Eds.), Handbook of Labor Economics. vol. 1. Elsevier, North Holland, pp. 641–692.
Rueda, David, Pontusson, Jonas, 2000. Wage inequality and varieties of capitalism. World Polit. 52 (3), 350–383.
Saint-Paul, Gilles, 2000. The Political Economy of Labour Market Institutions. Oxford University Press, Oxford.
Salverda, Wiemer, 2008. Labor market institutions, low-wage work, and job quality, In: Salverda, Weimer, van Klaveren, Maarten, van der Meer, Marc (Eds.), Low-Wage Work in the Netherlands. Russell Sage, New York, pp. 63–131 (Chapter 3).
Salverda, Wiemer, 2009. The bite and effects of wage bargaining in the Netherlands 1995–2005. In: Keune, Maarten, Galgóczi, Béla (Eds.), Wages and Wage Bargaining in Europe; Developments Since the Mid-1990s. ETUI, Brussels, pp. 225–254.
Salverda, Wiemer, 2013. The Evolution of Dutch Top-Income Shares Until 2012 and the Puzzle of Stability, AIAS Working Paper.
Salverda, Wiemer, Atkinson, Anthony B., 2007. Top incomes in the Netherlands over the twentieth century, In: Atkinson, Piketty, (Eds.), (Chapter 10).

Salverda, Wiemer, Haas, Christina, 2014. Earnings, employment and income inequality, In: Salverda, Wiemer, Nolan, Brian, Checchi, Daniele, Marx, Ive, McKnight, Abigail, Gyögy Tóth, István, van de Werfhorst, Herman (Eds.), Changing Inequalities in Rich Countries: Analytical and Comparative Perspectives. Oxford University Press, Oxford (Chapter 3).

Salverda, Wiemer, Haas, Christina, de Graaf-Zijl, Marloes, Lancee, Bram, Notten, Natascha, Ooms, Tahnee, 2013. Growing Inequalities and Their Impacts in The Netherlands, Country Report for The Netherlands. GINI project, http://www.gini-research.org/system/uploads/512/original/Netherlands.pdf?1380138293.

Salverda, Wiemer, Nolan, Brian, Maitre, Bertrand, Mühlau, Peter, 2001. Benchmarking Low-Wage and High-Wage Employment in Europe and the United States, Report to the European Commission DG Employment and Social Affairs. http://www.uva-aias.net/uploaded_files/regular/draftdef0-1-1.pdf Brussels.

Scheve, Kenneth, Stasavage, David, 2009. Institutions, partisanship, and inequality in the long run. World Polit. 61 (2), 215–253.

Schivardi, Fabiano, Torrini, Roberto, 2008. Identifying the effects of firing restrictions through size-contingent differences in regulation. Labor Econ. 15 (3), 482–511.

Schmitt, John, 2013. Why Does the Minimum Wage Have No Discernible Effect on Employment? Center for Economic and Policy Research, CEPR, Washington, DC.

Schwartz, Christine, 2010. Earnings inequality and the changing association between spouses' earnings. Am. J. Sociol. 115 (5), 1524–1557.

Shorrocks, Anthony, 1982. Inequality decomposition by factor components. Econometrica 50 (1), 193–211.

Shorrocks, Anthony, 1983. The impact of income components on the distribution of family incomes. Q. J. Econ. 98 (2), 311–326.

Slonimczyk, Fabián, 2013. Earnings inequality and skill mismatch. J. Econ. Inequal. 11 (2), 163–194.

Slonimczyk, Fabián, Skott, Peter, 2012. Employment and distribution effects of the minimum wage. J. Econ. Behav. Organ. 84 (1), 245–264.

Spitz-Oener, Alexandra, 2006. Technical change, job tasks, and rising educational demands: looking outside the wage structure. J. Labor Econ. 24 (2), 235–270.

Stewart, Mark, 2012. Wage inequality, minimum wage effects, and spillovers. Oxf. Econ. Pap. 64, 616–634.

Sutch, Richard, 2010. The Unexpected Long-Run Impact of the Minimum Wage: An Educational Cascade, NBER Working Paper 16355.

Thévenon, Olivier, 2013. Drivers of Female Labour Force Participation in the OECD, OECD Social, Employment and Migration Working Paper 145, Paris.

Thévenon, Oliver, Solaz, Anne, 2013. Labour Market Effects of Parental Leave Policies in OECD Countries, OECD Social, Employment and Migration Working Paper 141, Paris.

Van der Wiel, Karen, 2010. Better protected, better paid: evidence on how employment protection affects wages. Labor Econ. 17, 16–26.

Van Praag, B., Hagenaars, Aldi, van Weeren, Hans, 1982. Poverty in Europe. Rev. Income Wealth 28 (3), 345–359.

Van Weeren, Hans, Van Praag, Bernard M.S., 1983. The inequality of actual incomes and earning capacities between households in Europe. Eur. Econ. Rev. 24, 239–256.

Večerník, Jiří, 2010. Earnings Disparities and Income Inequality in CEE Countries: An Analysis of Development and Relationships, Working Paper 540, Luxembourg Income Study.

Večerník, Jiří, 2013. The changing role of education in the distribution of earnings and household income. The Czech Republic, 1988–2009. Econ. Transit. 21 (1), 111–133.

Visser, Jelle, 2009. The ICTWSS Database: Database on Institutional Characteristics of Trade Unions, Wage Setting, State Intervention and Social Pacts in 34 Countries Between 1960 and 2007. http://www.uva-aias.net/208.

Visser, Jelle, Checchi, Daniele, 2009. Inequality and the labour market: unions, In: Salverda, Wiemer, Nolan, Brian, Smeeding, Timothy (Eds.), The Oxford Handbook of Economic Inequality. Oxford University Press, Oxford, pp. 230–256 (Chapter 10).

Vlasblom, Jan Dirk, de Gijsel, Peter, Siegers, Jacques, 2001. Taxes, female labour supply and household income: differences between the Netherlands and the Federal Republic of Germany. Appl. Econ. 33, 735–744.

Vroman, Wayne, 2007. Replacement Rates and UC Benefit Generosity, Mimeo.

Wallerstein, Michael, 1999. Wage-setting institutions and pay inequality in advanced industrial societies. Am. J. Polit. Sci. 43 (3), 649–680.

Wood, Adrian, 1995. North–South Trade, Employment and Inequality: Changing Fortunes in a Skill-Driven World. Oxford University Press, Oxford.

World Bank, 2012. World Development Report 2013: Jobs, Washington.

Yun, Myeong-Su., 2006. Earnings inequality in USA, 1969–99: comparing inequality using earnings equations. Rev. Income Wealth 52 (1), 127–144.

CHAPTER 19

Cross-Country Evidence of the Multiple Causes of Inequality Changes in the OECD Area

Michael F. Förster*, István György Tóth†
*OECD, Paris, France
†Tárki Social Research Institute, Budapest, Hungary

Contents

Handbook of Income Distribution, Volume 2B
ISSN 1574-0056, http://dx.doi.org/10.1016/B978-0-444-59429-7.00020-0

Abstract

This chapter provides a thorough survey of what recent international (i.e., cross-country) studies can tell us about the multiple causes of income inequality in the OECD area with regard to both levels and trends. The survey covers economics literature in particular but also relevant evidence from sociology and political science. We provide an overview of drivers of inequality in six areas: (i) structural macroeconomic sectoral changes, (ii) globalization and technology change, (iii) labor market and other relevant institutions, (iv) politics and political processes, (v) tax/transfer schemes, and (vi) demographic and other microstructural changes. We find that the literature, while extremely rich in partial analysis of all six areas, provides very few analyses with truly multivariate and multicountry specifications for the joint section of the OECD and EU countries. Suggestions include more cross-discipline reflections on various findings. This is now well facilitated by the spectacular development of data, as well as in relation to methodological harmonization across disciplines.

Keywords

Income distribution, Globalization, Labor market institutions, Political economy, Redistribution, Demographic structure, Multivariate models, Cross-country comparisons, OECD countries

JEL Classification Codes

D30, D31, D63, I32, I38, J31, O15

19.1. INTRODUCTION

In their review of income inequality in richer and OECD countries, Brandolini and Smeeding (2009) concluded that "attempts to model and understand causal factors and explanations for differences in level and trend in income inequality across nations is the ultimate challenge to which researchers on inequality should all aspire" (p. 97). This sentence summarizes well the aim of the literature review in this chapter.

The chapter aims to provide a thorough survey of what international (i.e., cross-country) studies can tell us about the drivers and underlying causes of income inequality with regard to levels and, in particular, trends. The survey intends to be interdisciplinary, focusing on economics literature in particular but also on relevant evidence from sociology and political science.[1] While the overview intends to be comprehensive, some important research decisions limit its scope with regard to coverage and focus:

- The geographical coverage of the chapter is limited to the joint set of OECD and EU countries. Driving factors of inequality in emerging and developing countries and issues of world development are covered by Chapters 9, 11 and 20 in this volume.
- The chapter provides an update of existing reviews of literature with mostly recent studies, focusing largely on cross-country analyses that became available since the turn of the century.
- The chapter basically provides a meta-analysis based on review of the relevant literature. It does not produce a new data analysis within the frame of this survey. However, the chapter presents and provides a numerical analysis of the key findings of the literature.
- The focus of the chapter is on inequality of outcomes rather than inequality of opportunity. The analysis of the latter is provided in Chapter 4.
- Research results on determinants of poverty are not reviewed here. While it is acknowledged that (relative) poverty is a feature of inequality, we keep the focus here to studies aiming to explore the determinants of the full range of the dispersion of incomes. On poverty literature, see Nolan and Marx 2009, and Chapters 3, 8, 9, and 23 in this volume.
- When dealing with "inequality," the emphasis is on inequality of household income as much as possible, following the main focus of the Handbook. Given the scope of the empirical literature at hand, results of the determinants of the distribution of income subaggregates such as labor earnings also are reported. The determinants of the distribution of individual wages are, however, discussed in Chapter 18.
- The chapter focuses on the size distribution of personal incomes, leaving the vast range of literature on functional income distribution to other studies.
- While there is a trade-off between country coverage (N) and the length of the time series (T) in an analysis (given the limitations of data for large cross-country data sets for a long time series), the chapter draws practical boundaries here. A large cross-section of countries is relevant, even if only one or a few points in time are covered.

[1] The interdisciplinary approach applied here has forced us to make some difficult choices with regard to different methods and approaches applied by various strands of scientific analysis and that are rooted in the history of disciplinary accounts of inequality. Choosing as a starting point a frame that is (mostly) applied by economists might seem procrustean for representatives of other disciplines. With due acknowledgements, though, we hope our approach is useful.

On the other hand, analyses of only a few countries but for a long time series may be relevant for the review. The issue of this trade-off, however, is discussed further later in the chapter.
- The chapter reviews findings on the driving factors of inequality under several aspects: cross sections of within-country inequalities, quasi panels of countries and cross-country comparisons of longitudinal surveys (the data background of the studies is discussed in Section 19.3.2, covering the comprehensive data background of the income distribution literature). We do not include studies of cross-country differentials such as gross domestic product (GDP) convergence.

The structure of the chapter follows a broad classification of research questions of the literature. The chapter ends with a concluding section that attempts to summarize and classify the wealth of findings from the literature and to provide a critical assessment of the findings.

When selecting the empirical studies to be reviewed, we considered four elements as crucial: (i) the analyses had to show empirical results on income (or at least earnings) inequality; (ii) they had to cover a multiple of countries; (iii) they had to be at least multivariate; and (iv) their coverage had to relate to the joint set of OECD and EU countries. This led, obviously, to painful omissions of many excellent reports of driving factors of inequality.[2]

19.2. THE RESEARCH QUESTION AND METHODS TO EXPLAIN INEQUALITY AND ITS CHANGE

This chapter sets out the problem of a multicausal explanation of income inequality in a cross-country context. First we present the structure of the problem and then we provide an outline of the methods used in the literature we review.

19.2.1 The Structure of the Research Question

To understand and place the formulation of the research questions of the literature, it is useful to start with a very general flow chart showing the major elements of inequality formation (see Figure 19.1); this deliberately ignores potential causality directions at this stage. As the figure illustrates, income inequality (at all levels of economic development) is a product of macro processes (such as supply and demand processes, globalization, trade,

[2] However, these selection criteria could not always be fully respected. For example, the data background of certain studies we reviewed seemed at first glance to not properly fit the above criteria, for example, when a model of an important political process (such as corporatist agreements) is tested with individual wages rather than incomes. However, the line of argument dealing with the political economy of interest groups remains of interest even if it refers to the effect on wages only. Also, in some cases, especially in the frame of the debate on globalization and technological change, lessons from developing countries may be important for theoretical or methodological reasons, so some of those studies with coverage of countries outside of our prime target have not been excluded. The general guidelines from the above limitations, however, remain to be held.

Figure 19.1 Stylized description of determinants of income distribution.

and sectoral change in the economy); structural conditions (in terms of economic and social structures as well); and institutional constructs (political institutions for the aggregation of collective preferences, labor market institutions to assist an efficient utilization of human capital endowments, and tax/transfer schemes for institutionalizing redistribution in society).

Schematically, Figure 19.1 numerates six families of potential key drivers of earnings and income distributions. From left to right, "globalization" is primarily meant to cover the economic dimensions of globalization, such as increased trade integration, outsourcing or financial integration. Technological changes also fall into this family. Next, under the heading "labor relations and regulations," we also discuss institutional features of the labor markets, such as the level of unionization, the potential role of wage-bargaining institutions, or levels of corporatism embedded into the political system. "Political processes" include preference formation (of voters and of parties), political representation, and interest group politics. "Redistribution and tax-transfer policies" involve various policy arrangements aimed at altering the "original" distribution that came about as a result of market processes. "Demographic and societal structure" refers to the way individuals (with their own incomes) combine into families and households (household structure by age, employment, income levels) and how the society is composed of various sociodemographic variables (such as age, gender or education). Finally, the "macroeconomic structure" of societies (characterized by sector distribution of employment, by degrees of labor market attachment, etc.) is of central importance for the determination of overall inequalities. With this schematic we illustrate the complexity of factors that affect income distribution and highlight the partial nature of most empirical analyses we found in our literature search.[3]

The overwhelming majority of the articles we reviewed model inequality (or inequality changes) and regress a chosen inequality measure on selected driver variables, usually among one (rarely more) of the driver families. Among this literature, the list of 48 articles with key features analyzed and that come closest to satisfying the criteria above can be found in Annex Table A19.1. Many of them focus on some particular parts of Figure 19.1. Few of them, however, aim to cover the full range of potential variables explaining changes in income distribution (Cornia, 2012, or OECD, 2011, are among these exceptions; see Annex Table A19.1 for further details). Nevertheless, it is useful to keep the full picture in mind when certain specific parts are analyzed.

A general formulation of the approach taken can be written in the form of a generalized regression equation (Equation 19.1).

$$INEQ_{i,t} = \alpha + \beta * X_{i,t} + \gamma * Z_{i,t} + \lambda * Q_{i,t} + \eta_i + \mu_t + \varepsilon_{i,t} \tag{19.1}$$

where $INEQ_{i,t}$ is a properly chosen measure of inequality of household incomes within country i at a certain point in time t, and $X_{i,t} = \{x_{j,i,t}\}$ is the vector of population characteristics aggregated from individual (or household) attributes (age, education, sex, household type, etc.). On country level these attributes define the structural conditions to inequality development in a certain country. $Z_{i,t} = \{z_{j,i,t}\}$ is the vector of macroeconomic (GDP, trade, financial globalization, technology, etc.) and institutional variables (policies, redistribution, wage-setting mechanisms, etc.). In a cross-country comparison, where the unit of analysis is country, these variables enter as attributes of the macro units (countries); $Q_{i,t} = \{q_{j,i,t}\}$ is the vector of specific historic/contextual variables (history, size location, composition, etc). η_i And μ_t stand for the inclusion of country and time dummies, respectively (these occasionally entail, as fixed effects, a large variety of country-specific attributes and year-specific effects). $\varepsilon_{i,t}$ Represents the error term, i is $1, \cdots, N$ for countries, and t is $1, \cdots, T$ for years. For later use we denote Equation (19.1) as a grand inequality regression equation (GIRE).[4]

[3] By nature, our account of the literature—while covering a wide range of areas, as shown in Figure 19.1—remains superficial from a specialist point of view. We are, however, in a favourable position insofar as a number of chapters in this Handbook provide more in-depth detail for all six areas. For instance, although we discuss the effects of labor market institutions on income distribution, some particular elements such as wage policies are further detailed in Chapter 18. Similarly, while we include a discussion of the effects of redistribution and analysis of tax/benefit schemes, these are not exhaustive given that Chapters 24 and 25 are devoted to these issues. Further examples of complementarities could be listed.

[4] Although it looks very general, the way the equation is formulated here is, to some extent, also very specific. More refined formulations, of course, also have to take nonlinearities and potential interactions between explanatory variables into account. We, however, offer the formulation here as a heuristic device only, to help structure the frame for the chapter. Another caveat is that the implications for inequality depend on the specification of the left-hand side. This is discussed later in Section 19.4.

19.2.2 Notes on the Arguments and Parts of the Grand Inequality Regression Equation (GIRE)

Atkinson and Brandolini (2009) advise readers hoping to understand the empirical inequality literature that they should consider theory, data and estimation together, meaning that data have to be sufficient and adequate to theory, and estimation methods have to be adequate to available data. This requirement is key for the interpretations of empirical articles in all disciplines (economic, sociological and political science literature). When going through the various empirical accounts, we focus attention on this requirement.

19.2.2.1 The Usefulness of the General Formulation

An important point concerning the regression approach should be addressed at the outset. Some scholars may argue that cross-country regressions fail to capture adequately the cross-country differences because historical and institutional specificities define completely different relationships between dependent and independent variables. Others argue that the relationship between variable X and variable Y will be the same when controlled for all other potential factors. We think that well-specified regressions can help in understanding links (even if not causalities) between various factors, but, at the same time, caution is warranted, and country specificities always have to be taken into account. Classifications of various welfare regimes (going back to the seminal work of Esping-Andersen, 1990, differentiating between the conservative, the liberal and the socialist regimes) or differentiations between such complex settings as varieties of capitalism (Hall and Soskice, 2001) can add important parameters, and they do describe different sets of circumstances, but controlling for them (in an ideal data case) leaves sufficient room for the relationship between X and Y to operate uniformly across countries.[5]

Taking—admittedly—to the extreme the welfare regime literature, however, makes it quite difficult to identify the contribution of the various single factors to income distribution (or a change in it). Given that welfare regimes are defined as a complex interplay and a joint product of the state, the market and the family (Esping-Andersen, 1990; Esping-Andersen and Myles, 2009), the proper methodological analogy would be cluster analysis rather than regression. Clusters built from a wide array of country attributes could show similar and dissimilar country examples of inequality, together with the other observed factors (Kammer et al., 2012 is a prominent example of this type of analysis of welfare regimes). However, no causality directions could even be attempted. Without even hints to any judgments on this, we try to comply with the logic highlighted above to help structure the discussion of determinants of income distributions.

[5] In fields where institutional complexities of the subject and the training background of scholars induce widespread use of qualitative methods, an explicit mention of this caveat is important (see, for example, Rueda and Pontusson, 2000 warning for political scientists or Kenworthy, 2007 message to sociologists).

19.2.2.2 The Units of Analysis

In cross-country explanations of inequality drivers, the units of analysis (data points) are countries, characterized by various inequality measures as left-hand variables and other macro characteristics such as GDP, shares of economic sectors, globalization, institutions or redistribution as right-hand variables. In most of these analytic attempts a time dimension is introduced on the right-hand side with the use of multiple data points for various periods. This in some cases allows for a macro-level analysis of changes.[6] Many reviewed studies belong to this class. It would, however, be ideal to have analyses of pooled micro-data to identify cross-country differences of determinants of income inequality. Surprisingly enough, we did not find articles that fit into the latter category.

Another strand of analysis, again using micro rather than macro variables to explain the underlying drivers of inequality (and of changes in inequality), makes use of decomposition methods. Decomposition can be a powerful instrument to disentangle mathematically the different components that make up overall inequality. Decomposition can be used to identify the relative roles of several income sources to overall income inequality (tracing back to Shorrocks, 1982) or else to analyze the contribution of different population subgroups to levels of and trends in inequality.

19.2.2.3 Regression Methodology

The majority of macroeconomic cross-country panel studies reviewed use ordinary least square (OLS) regression with pooled cross sections in a macroeconomic setting to gauge causal factors impeding between- and within-country inequality. However, simple pooled OLS approaches have been judged unsatisfactory by many authors of multicountry studies of trends, especially if the analysis contains a larger sample of countries that differ in a systemic way—either in measuring inequality or in institutional or macroeconomic specificities. For example, there may be unobserved time-invariant, country-specific heterogeneity that forces an error term relating to a same country over time being correlated, leading to biased estimates of traditional OLS methods. Moreover, there may be panel heteroscedasticity because (i) error variances for a given country may display time dependence (i.e., serial correlation) and/or (ii) error variances may systematically

[6] When change in country level inequality indexes is of interest, using the notations of Equation (19.1), a the following relationship is estimated:

$$\delta(INEQi) = f(\delta Xi, + \delta Zi, + Qi, + \eta i, + \varepsilon i), \tag{19.2}$$

which should be read that change in inequality (on a country level) is dependent on a specifically weighted portfolio of the following factors:

δX = change in structural attributes (age, education, sex, etc.) from t to $t+1$; and

δZ = change in macro and institutional variables (policies, redistribution, wage-setting mechanisms, etc.) from t to $t+1$.

The other arguments remain the same as in Equation (19.1).

differ across countries. Both patterns would lead to inefficient OLS estimates if not treated properly.

To assign country-specific factors to country-specific intercepts rather than constraining all countries to the same intercept, a large majority of the macroeconomic panel approaches reviewed here apply fixed effects in their models. Gourdon et al. (2008), for instance, put forward as one of their main conclusions that "results from studies that do not control for effects of omitted variables via fixed effects are biased" (p. 352).

However, some authors consider fixed-effects methodology overly conservative because any variation between countries is disregarded in the data and the effect of some factors that are constant over time but differ between countries, such as institutions, are likely to be overlooked. This is the line of argument of, for example, Nielsen and Alderson (1995) and Alderson and Nielsen (2002), who propose as an alternative a random effects model ("random" in the sense that it treats unobserved effects as random variables because they are treated independent of the explanatory variables). Such a model removes only a fraction of the country-specific means, not the whole mean, and is thus considered as "less wasteful of between-country variation" (Alderson and Nielsen, 2002, p. 26).

There also has been more general criticism of the usefulness of time series regression methodology for explaining inequality determinants. One issue is that of identifying long-running relationships and cointegration of series. A problem with the standard panel regression approach is how to account for the timing of the effect of the explanatory variables. Globalization or deregulation, for instance, may well be "significant" factors but they may take some time to affect the distribution; furthermore, the delay may not be the same across countries and across factors. This may be less of a problem if long-enough time series were available, but this is generally not the case.

A related issue is that of the nonstationarity of data points, that is, that they have means and variances that change over time, either in trends, cycles or at random. Parker (2000), for instance, argues that the fact that many explanatory variables are likely to be nonstationary produces spurious regression results in that they may indicate a relationship between two variables where there is none. Further, the power of integration and cointegration tests tends to be low when small sample sizes are used, which is often the case in studies of inequality.[7] One possible solution is to combine OLS with the method of error correction models proposed by Hox (2002) and applied, for instance, by Rohrbach (2009) or Cassette et al. (2012). This method regresses the lowest-level variable on covariates from all other levels simultaneously.

Similarly, Jäntti and Jenkins (2010) argue that direct estimation of parameters in time series analysis can be problematic because of the nonstationarity of both left- and

[7] As an alternative going beyond the OLS approach, Parker (2000) proposes turning to decomposition and cross section regression analyses.

right-hand variables. Further, left-hand variables are typically bounded, usually to the unit interval, which involves problems for tests of stationarity and also raises more general issues about the appropriate specification.[8] Jäntti and Jenkins (2010) propose returning to parametric distribution functions instead. Applying the latter approach to UK data, they found a lesser distributive effect of macroeconomic factors than is suggested when commonly used methods are applied.

That said, even if cross-country panel regressions entail a number of interpretational problems and often, taken together, provide inconclusive findings, especially with regard to the role of globalization, much has been learned from the studies undertaken during the past decades. As Eberhardt and Teal (2009) put it (referring to controversial findings of cross-country growth regressions), "the lesson of incomplete success is not to abandon the "quest" but to seek to understand why success has been so limited" (p. 28).

The most common approach to explaining changes in inequality in the studies reviewed is with aggregate inequality measures. By doing so, however, one might miss important changes in the distribution. From that point of view, it may be worth pursuing more comprehensive approaches, such as the reweighting procedure proposed by DiNardo et al. (1996), as well as the recentered influence function regressions by Firpo et al. (2009) for labor market analyses or the microeconometric approach by Bourguignon et al. (2005) to the household income distribution in the microeconomics of income distribution dynamics project. All these approaches aim to shed light on the drivers behind changing income distributions by simulating counterfactual distributions in a controlled manner.

Such approaches remain on a partial equilibrium view. Another challenge today, therefore, is to bring together macro- and micro-based regression methodologies and their findings. To that aim, new tools of macro–micro models have been developed (see Bourguignon et al., 2010). These models analyze, for example, the distributive effect of "macro" events, such as migration, by integrating a macro framework with a microsimulation model that uses household or individual data, either by implementing a sequential approach (e.g., first computing the macroeconomic variables in a computable general equilibrium model and then using estimated values as input for a microsimulation model that distributes the effects of macro changes among micro units), or via full integration of microsimulation models within computable general equilibrium models.

In terms of the presentation of results from cross-country panel regression studies, in addition to indicating the significance of coefficients, many studies try to gauge the *relative* importance of the different variables that have been estimated to affect inequality. Because the variables under examination often are measured in different units, a common approach is to calculate standardised coefficients (which are obtained by first

[8] Following Atkinson et al. (1989, p. 324–325), there is a case for using a log-logistic formulation of the type $\log[INEQ/(1-INEQ)]$, which allows unbounded variation.

standardizing all variables to have a mean of 0 and a standard deviation of 1). Moreover, simple simulations or a back-of-the-envelope calculation often are used to quantify the effect of an individual factor. For instance, IMF (2007) and Jaumotte et al. (2008) calculated the contributions of various factors to the change in inequality as the annual average change in the respective variable multiplied by the corresponding coefficient, and the averages across country groups were weighted by the number of years available for each country (to increase the weight of countries with longer observation periods in these averages). The OECD (2011) makes use of the same computation approach to show the relative size of the contributions of different factors to the increase in overall earnings inequality.

19.3. DATA SOURCES FOR CROSS-COUNTRY STUDIES

This section provides an overview of data available for multivariate analysis of within-country inequality in an international comparison.

19.3.1 Different Strategies for Multicountry Studies

At the outset, although seldom explicitly, research needs to decide on the precise coverage of a country sample to be analyzed. While this choice may be constrained (but should not be motivated) by the availability of data, two different strategies exist when using multicountry samples to explain variations in inequality. First, the sample may be formed by a set of countries sharing similar systemic characteristics (e.g., the OECD area), a strategy called "most similar design" by Przeworski and Teune (1970). Conversely, the aim can be to test a hypothesis such as the Kuznets-type relationship between development and inequality on a set of countries with a maximum of differing systemic characteristics, a "most dissimilar design" strategy.

While many earlier studies of global causes of inequality aimed to include as many countries as possible to the analyses, they still had an overrepresentation of developed countries in the sample. Coverage of African countries in particular was very low. In a typical study with "universal" coverage of inequality observations in the 1990s and early 2000s, OECD countries represented half up to two-thirds of the whole data set. This has changed in more recent studies, but the OECD area still makes up typically a third of all country observations. While this choice is dictated by data availability, the precision and generalization in the interpretation of empirical results suffers.[9] Depending on the nature of a research question and following a thorough examination of underlying data and their quality, a reduced sample of countries may be a preferred option, or, as Atkinson and

[9] A good example is the discussion by Tsai et al. (2012), who replicate the same model on the same data as Zhou et al. (2011) but find different and partly contradictory results by adding dummy variables for developed, transitional and developing countries rather than pooling all 60 countries included in the study.

Brandolini (2006a) propose, "A deeper understanding of national sources ⋯ may lead us to analyze a carefully matched subset of countries, rather than to seek to maximize their number."

In that sense, even the focus on an apparently more homogenous country panel such as the group of OECD or EU countries may involve interpretational problems, especially if new member countries are included in the analysis. The results from empirical analyses of the importance of sector dualism and sector bias between agriculture and industry as a driver of inequality (see Section 19.3.2) very much depends on how the OECD area and the EU area are defined. Empirical findings may be blurred if these definitions include not only the "traditional" OECD member countries or the "old" EU member states but also newer member countries such as Poland or Mexico, where the share of agricultural employment is still important (more than half the OECD average) and the dualism model may have some salience. In what follows we go through the "menu" of the available data sets for inequality research.

19.3.2 Data Sources: The Fast Development of Data Availability in the Last Decades

In the concluding remarks of his seminal article on economic growth and income inequality, Simon Kuznets (1955) acknowledged that his "paper is perhaps 5 per cent empirical information and 95 per cent speculation" (p. 26). Until the early 1990s, the availability of internationally comparable income inequality data still was scarce. During the past two to three decades, however, a substantive amount of household surveys became available, and much progress in distributional data collection and standardization has been made in OECD countries. The situation is still far from being ideal, but today's research and results may perhaps mirror 50% empirical information and 50% speculation.

This section is about the former 50%. It describes the main sources of data on income inequality and other key variables used in cross-country studies of the drivers of inequality. It reviews international data sets of income inequality: *ex ante* standardized data, ex post standardized data, data standardized on best national sources, and secondary data sets. The review focuses on data sets that include at least most of the group of OECD countries. It will also become clear how some of these new data sources open prospects for new types of research questions and application of new types of analytic methods (notably the use of longitudinal panel data).

19.3.2.1 Standardized Microdata

Despite continuing progress, the availability of comparable primary data sets for inequality research is still limited. The major initial and pioneering effort was launched 30 years ago by the data collection of the Luxembourg Income Study (LIS). Since the mid-2000s, the Statistical Office of the European Union (EUROSTAT) launched a harmonized household survey on income and living conditions (EU-SILC), which is available for the 28 member countries and some additional European countries. Equally, since the mid-2000s, OECD has made available a detailed set of standardized household income and poverty indicators for their 34 member countries.

19.3.2.1.1 Luxembourg Income Study

The LIS, formerly known as the Luxembourg Income Study, is a data archive and research centre dedicated to cross-national analysis (http://www.lisdatacenter.org/). The project collects income microdata from household surveys and standardizes those into a common framework of income, demographic and employment variables. The standardization is undertaken ex post. The key concept is that of disposable income, and detailed income aggregates are available. When the project started, it included data from seven countries. Today, LIS stores microdata for over 40 countries, for 8 points in time, starting with a year around 1980, in approximately 5-year intervals. Access to the LIS microdata is granted to researchers of financially contributing countries and institutions and students worldwide upon registration. Use of the microdata is permitted for scholarly, research or educational purposes but not for commercial purposes.

One of the key assets of the LIS database is that it allows researchers the access to the microdata, via a remote access system. The scrutiny of the ex post standardization also allows a high degree of comparability of the micro variables. One main disadvantage is the somewhat limited geographical and time coverage, although the recent inclusion of a number of middle-income and emerging countries as well as a more frequent update (3-year rather than 5-year intervals) will allow more extensive panel data analysis (http://www.lisdatacenter.org/our-data/lis-database/documentation/list-of-datasets/).

19.3.2.1.2 EU Statistics on Income and Living Conditions

The EU-SILC is an annual survey that collects microdata on income, poverty, social exclusion and living conditions in the 28 EU member countries and 4 non-EU countries. It has been implemented since 2004 for 15 countries and since 2007 for 32 countries (http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/eu_silc). The EU-SILC surveys are "output" rather than "input" standardized. This implies that the data are not collected with a single survey across all countries; rather, countries are provided a list of variables that they can collect using national surveys and definitions, and the necessary standardization is made on this basis by EUROSTAT. EU-SILC includes longitudinal information insofar as the surveys are based on a rotational panel (usually with a duration of 4 years). In contrast to most other longitudinal surveys, cross sectional and longitudinal data are released separately in the EU-SILC.

Access to the anonymized EU-SILC microdata (the so-called user database) is not granted to individuals but only to research institutions (or similar entities) inside the EU and European economic area countries by means of research contracts. For other kinds of organizations inside the EU and organizations outside the EU, approval for access needs to be requested from the European Statistical System Committee, which takes about 6 months. A detailed set of indicators on incomes and other living conditions from these data is available from the EUROSTAT databank (http://epp.eurostat.ec.europa.eu/portal/page/portal/income_social_inclusion_living_conditions/data/database).

The main assets of the EU-SILC are the high degree of standardization, especially with regard to income concepts; the availability of annual data; and the availability of a longitudinal part of the data. One disadvantage for researchers today is simply the fact that the project is still relatively young: microdata are generally available for less than 10 years, thus preventing the analysis of long-term series.[10] There are also a few remaining problems that have to do with the loss of some information when the wealth of original microdata is transformed into a more restricted final data set for which the underlying methodology of such transformations as well as treatment of data at the national level (e.g., imputation procedures) are not always exhaustively documented. That said, it has been suggested that most of the latter set of problems can be easily overcome with a greater consistency and clarity in documentation in the years to come (Iacovou et al., 2012).

19.3.2.1.3 OECD Data (Income Distribution Database)

The OECD income distribution database (IDD) builds on regular data collection undertaken by the OECD through a network of national consultants who provide standard tabulations from national microdata considered the "most appropriate" data source in each country and are based on comparable definitions and methodological approaches. This is done via a detailed data questionnaire consisting of tabulations on income distribution and poverty indicators, together with standardized terms of references. The main concept of the data collection is that of equivalized household disposable income, including wages and salaries, self-employment incomes, realized property incomes and cash transfers from the general government less taxes and social security contributions paid by households. The definitions used in calculating these income components are based on the recommendations for household income statistics adopted by the Canberra Group (see http://www.unece.org/stats/groups/cgh.html).

A detailed set of variables for the 34 OECD member countries is available from the OECD "data cube" (http://www.oecd.org/social/income-distribution-database.htm). It includes several summary inequality and poverty measures (on a before and after tax/transfer basis) as well as data on income levels and population ventilations. Data

[10] Doubts as to the comparability of EU-SILC with a predecessor survey, the European Community Household Panel, which covered 15 EU countries for the years 1994 to 2001, remain. It should also be noted that the current practice of EUROSTAT publications is to report the EU-SILC survey year n for indicators but not the income year, which is $n-1$ in all countries except Ireland and the United Kingdom. This can create confusion when comparing EU-SILC-based indicators with results from other surveys.

are available in approximately 5-year intervals back to the mid-1990s and, for a subset of countries, to the mid-1980s and mid-1970s. From the mid-2000s, data are available on a more frequent basis, depending on the underlying surveys but, in general, annually (for 28 of the 34 countries). Access to these data is free.

The method of data collection used by the OECD IDD allows coverage of the entire region of OECD countries with harmonized data that facilitate cross-country comparison, based on information that is both more up to date relative to that available through other statistical sources and better suited for assessing changes in income distribution over time. However, data are available only on an "equivalized" household basis, which renders comparison with indicators on a "per capita" basis (used in many of the more global data sets) very difficult. The main disadvantage of the OECD database is that it does not allow access to the original microdata, which constrains the analyses that can be performed. In that sense, the OECD income distribution database constitutes its own category between primary and secondary data sets.

19.3.2.2 Secondary Datasets

With regard to the difference of the data sets described above, secondary datasets are based on a collection of published or otherwise available summary key inequality indicators. These usually include the Gini coefficient, quintile share ratios and/or percentile ratios and, more rarely, other summary measures such as the Theil index. Often, alternative series for the same country and year point are proposed alongside recommendations of "preferred" series, along the lines of, for instance, the A–B–C typology used by Atkinson (2008).[11] Typically, such data sets aim to collect indicators for the greatest number of countries. The trade-off is that there is necessarily less room available for verification of data quality and consistency, which leads to issues of data comparability between and within countries.

19.3.2.2.1 The Deininger-Squire Data Set (Measuring Income Inequality Database)

Klaus Deininger and Lyn Squire brought together a large set of worldwide inequality indicators in 1996. Their data set (DS) compiled Gini coefficients and cumulative quintile shares for 138 developed and developing countries, adding summary information on the nature of the data (population coverage, income or consumption base, net or gross income base). Most of the data cover the period between the 1960s and early 1990s.

With regard to earlier data compilations,[12] the DS data set imposed "minimum standards for quality," namely that indicators are based on household surveys, on

[11] Atkinson (2008) undertakes an in-depth review of available data sources on earnings inequality and classifies them into three groups: (A) most appropriate, (B) acceptable if not ideal and (C) rejected.

[12] In the early 1970s, the first major improvements of international data comparisons were achieved by Jain (1975) and when Adelman and Morris (1973) and Paukert (1973) tested the Kuznets hypothesis.

comprehensive coverage of the population, and on comprehensive coverage of income sources (Deininger and Squire, 1996, p. 567). On this basis, among the entire data set of 2630 observations, Deininger and Squire identify a subset of "high-quality" observations, with 693 observations for 115 countries. Those observations labelled "accept" in the DS data set nonetheless include indicators based on different definitions and methodologies, which impedes the comparability of these data.[13]

The DS data set is freely available at the World Bank's website (http://go.worldbank.org/UVPO9KSJJ0). It became a major data source for international inequality research during the early 2000s, including many of the cross-country panel studies reviewed in this chapter. While there were further developments on the basis of the DS data sets in the frame of follow-up projects (see 19.3.2.2.2), the above-mentioned version has not been updated or revised for corrections.

19.3.2.2.2 UNU-WIDER Database

In the vein of the DS data set and partly based on it, the United Nations University-World Institute for Development Research (WIDER) World Income Inequality Database (WIID) collects a secondary inequality data set for developed, developing and transition countries. The project started in the late 1990s and led to a first release of data for 155 countries (WIID1), extending the time frame to the early 2000s and augmenting the number of distributional indicators: calculated and reported Gini coefficients, decile and quintile shares, as well as survey means and medians, along with the income shares of the richest 5% and the poorest 5%. In addition to income and consumption, the data set also includes indicators for earnings.

A second and substantially revised version of WIID was compiled in the mid-2000s and resulted in the release of WIID2. The currently available version—World Income Inequality Database V2.0c (May 2008)—proposes data series up to 2006 and is described by the authors as a "new" rather than "updated" data set. It adds, where possible, a second Gini coefficient estimate calculated using a method developed by Shorrocks and Wan (2008) to estimate the Gini coefficient from decile data. An update of the database to WIID3.0 is pending at the time of writing.

[13] Deininger and Squire accept both person- and household-based Gini coefficients because the mean difference between these estimates turned out to be not too large (<2 points), and they therefore do not expect a large systematic bias in empirical work. A similar argument leads to the inclusion of both gross income- and net income-based indicators, with an average difference of 3 points found in 19 developed countries, and on the grounds that redistribution is more limited in developing countries. The DS data set also includes both income- and consumption-based indicators because 39 countries (136 observations) report only the latter. Because the bias can be larger in this case, one suggestion was to add the mean difference of 6.6 points found between the expenditure-based and income-based coefficient to the former (Deininger and Squire, 1996, p. 582).

Similar to the DS data set, WIID defines three quality criteria—(i) whether the underlying concepts are known, (ii) coverage of concepts and (iii) survey quality—but provides a more detailed quality ranking from 1 (underlying concepts are known and the quality of the income concept and survey can be judged as sufficient) to 4 (unreliable).[14]

The WIID dataset is freely available at the UNU/WIDER website (http://www.wider.unu.edu/research/Database/). It has been increasingly used in international inequality research and, with the merge with the former DS data set, constitutes the most widely known secondary inequality data set. One of four articles reviewed in Annex Table A19.1 make use of this data set.

19.3.2.2.3 All the Ginis Data Set

The All the Ginis (ATG) data set has been put together by Branko Milanovic from the World Bank since 2004. It includes combined and harmonized Gini coefficients (but no further inequality indicators) from seven original sources: the LIS, the Socio-Economic Database for Latin America, the EU-SILC, the World Bank Europe and Central Asia data set, the World Income Distribution (WYD), World Bank PovCal, and the WIDER.

The most recent version of the ATG data set was released in 2013 and includes close to 4000 Gini observations for 164 countries for the period from 1950 through 2012. Almost 2000 of these observations have been considered "consistent." Rather than classifying observations as "accept" (DS) or "reliable" (WIID), this "consistent" classification is based on an approach described as "choice by precedence." This approach takes the Gini values in overlapping cases in order of preference of the seven data sources, namely in the order as they are listed above.[15] The ATG data set presents the Gini values along with key dummy variables defining the type of welfare aggregate (income or expenditure, net or gross) and recipient unit (household or individual). Another specific feature of the ATD data set is that it includes a variable that allows the survey to be distinguished from the income year.

The ATG data set is freely available in form of a stata file at http://econ.worldbank.org/projects/inequality.

19.3.2.2.4 WYD Data Set (World Bank)

The WYD database was created as part of the World Bank's work on global income distribution. The objective of this work is to gather and analyze detailed household survey data for as many countries as possible for several benchmark years to calculate estimates of global inequality. Currently, data exist for five benchmark years (1988, 1993, 1998, 2002

[14] Regarding the difference of the DS data set, there can be more than one observation labelled 1 for the same country and the same year. In some cases there can be up to six observations with label 1, such as in the case of Germany for 1984.

[15] The database allows the user to define any alternative "choice by precedence."

and 2005). The objective of the WYD database was to create "rich" (numerous in terms of countries) and "dense" (ventiles or percentiles for each country's distribution) coverage for the benchmark years, not to maximize the number of Gini observations or provide longer-term series for individual countries. The WYD series are integrated into the ATG data set described earlier.

The WYD data are freely available in form of a stata file at http://go.worldbank.org/IVEJIU0FJ0.

19.3.2.2.5 The PovCal Database (World Bank)

The PovCal database covers the period since 1978 and includes 124 low-income, lower-middle-income and upper-middle-income countries, thus excluding higher-income OECD countries. In general, PovCal shares the same underlying survey data sources as WYD. There are over 800 Gini observations, most of which are calculated from direct access to household surveys. The PovCalNet tool is available at http://iresearch.worldbank.org/PovcalNet/index.htm.

19.3.2.2.6 World Development Indicators (World Bank)

The World Development Indicators (WDI) is the primary World Bank collection of development indicators compiled from officially recognized international sources. These also include the Gini index. However, data on OECD countries are scarce, with many countries missing data in all years. A priori, WDI Ginis also should come from the same underlying microdata used by WYD and PovCal. The data are available at http://databank.worldbank.org/data/home.aspx.

19.3.2.2.7 Sociómetro-BID (Inter-American Development Bank)

Sociómetro-BID is a diverse data set of social indicators derived from national household survey data, covering 21 Latin American and Caribbean countries from 1990 to 2009. While the Sociómetro includes traditional global indicators including the millennium development goals, the database also includes information on Gini coefficients for per capita household income. The data are freely available at http://www.iadb.org/research/sociometroBID/tables.cfm?indicator=4&lang=en.

19.3.2.2.8 TRANS-MONEE Database (UNICEF)

The TransMonEE (Transformative Monitoring for Enhanced Equity) database collects a vast range of data relevant to social and economic issues in 28 countries of Central Eastern Europe and the Commonwealth of Independent States. The database was initiated by the UNICEF Innocenti Research Centre in 1992 and is updated annually. The 2012 version of the database contains 180 economic and social indicators divided into 10 topics (population, natality, child and maternal mortality, life expectancy and adult mortality, family formation, health, education, child protection, crime and juvenile justice, economy).

It includes data on Gini coefficients, covering the period 1989–2009. In general, these data are based on interpolated distributions from grouped data from household budget surveys. The data are freely available at http://www.transmonee.org/.

19.3.2.2.9 International Labor Organization Database

Since 2012, the International Labor Organization (ILO) database provides recent data for over 100 indicators and 230 countries. It includes a series of D9/D1 and D9/D5 percentile ratios for earnings for employees (although the precise definition and concept are not clear from the description). The data are freely available from http://www.ilo.org/ilostat/faces/home/statisticaldata. The former ILO database LABORSTA (http://laborsta.ilo.org/) included both decile values and Gini coefficients for selected years up to the early 2000s.

19.3.2.2.10 The GINI Inequality and Poverty Dataset

The GINI Inequality and Poverty Dataset is a very recent outcome of the "Growing Inequalities' Impacts" (GINI) project completed within the 7th Framework program of the European Commission between 2009 and 2013. The project produced in-depth case studies for the 30 participant countries, which include 25 of the 28 EU countries together with 5 non-European countries: Australia, Canada, Japan, Korea and the United States. The country case studies followed a predetermined template specifying the most important variables to be monitored over a 30-year time span (from 1980 to 2010). The variables related to inequality cover Gini coefficients and relative income poverty. For both Gini coefficients and poverty, the preferred income concept is net/disposable equivalized household income. The income sharing unit is the household, whereas the unit of analysis for the computation of various indexes is the individual member of the household. In each case the figures refer to national coverage and thresholds rather than, for example, regions or specific social groups. For most of the countries and for most of the data points these requirements are met (for further details see Salverda et al., 2014 and Tóth, 2014).

19.3.2.2.11 Chartbook of Economic Inequality Data

Atkinson and Morelli (2014) created a chartbook of economic inequality that includes indicators beyond income inequality measures for 25 countries (of which 17 are OECD countries) and covers series for up to 100 years until the present. These refer to earnings inequality (usually D9/D5 ratios for OECD countries) and overall inequality (usually Gini coefficients of household income) as well as poverty, pretax top income shares and wealth. These series are based on "preferred" definitions, which are documented for each country included in the data. The focus of this data collection is on over-time comparability rather than between-country comparability. The underlying data are freely available at www.chartbookofeconomicinequality.com.

19.3.2.2.12 World Top Incomes Database

Long-run data series on pretax top incomes ranging back 80 years or more have been collected and prepared by Facundo Alvaredo, Tony Atkinson, Thomas Piketty, Emmanuel Saez and various collaborators and are available online (http://topincomes.parisschoolofeconomics.eu/). The database includes information on top income levels and top income shares (such as the top 1%, top 0.1% or top 0.01%) for 27 countries, of which 18 are OECD countries.

Two main limitations of these data sets are that they cannot be used to describe the whole distribution (and hence do not include summary inequality measures) and that data refer to pretax incomes. Further limitations of tax data for inequality analysis are that tax-exempt income is typically not reported and consequently is left out of the indicators; cross-country differences (and changes over time) in the concept of income that is measured; the extent of tax planning and tax evasion; and the definition of the tax unit. For a summary of the main results from analyses of these data and a discussion of the underlying data, see, for instance, Atkinson et al. (2011).

19.3.2.3 Secondary Synthetic Data Compilations

Synthetic data compilations are based on regression-based procedures to estimate time series from existing inequality data sets.

19.3.2.3.1 University of Texas Inequality Project

The University of Texas Inequality Project data set, which is associated with the work of James Galbraith, is based on a project concerned with measuring and explaining movements of inequality in wages and earnings and patterns of industrial change around the world. It uses microdata available based on industrial statistics from the United Nations Industrial Development Organization. The project establishes a relationship between these measures and the broader concepts of inequality, such as income inequality, which is considered reasonably reliable. The data use the Theil's T statistic to compute inequality indexes from industrial, regional and sectoral data. It produces data sets on pay inequality at the global level; at the national level, including data for Argentina, Brazil, Cuba, China, India, and Russia; and at the regional level for Europe. Data on pay inequality were used as an instrument to estimate measures of household income inequality for a large panel of countries from 1963 through 2008. This global data set has around 4000 country-year observations. All data sets are available at http://utip.gov.utexas.edu/data.html.

19.3.2.3.2 SWIID Database

The SWIID database standardizes the WIDER data (described earlier) and other inequality data while minimizing reliance on problematic assumptions by using as much information as possible from proximate years within the same country.[16] It uses the data collected by the LIS as the benchmark standard. The SWIID currently incorporates Gini indexes of gross and net income inequality for 173 countries for as many years as possible from 1960 to the present, as well as estimates standard errors for these statistics. The SWIID data and the procedure used to generate it are available at http://myweb.uiowa.edu/fsolt/swiid/swiid.html and are described by Solt (2009).

There are other, more one-off exercises to build synthetic cross-country data compilations from existing inequality data sets, such as the Standardized Income Distribution Database. This database was created by Babones and Alvarez-Rivadulla (2007) on the basis of the UNU-WIDER dataset (WIID) but is not available online. It can be requested from the authors.

19.3.3 Concluding Remarks

There is no single "ideal" data set for international research on the multiple causes of inequality, despite the rapid development of international data sets of primary and secondary inequality data over the past 20 years. Opting for one or the other of the above-described data sets depends on the nature of the research question as well as on the target group of countries that are to be compared. If a study is confined to the group of EU and/or OECD countries, one of the primary data sets may reveal the first choice because of their higher degree of standardization. For more global country coverage, secondary data sets provide a necessary starting point but great care needs to be taken, and not all series can be integrated in econometric analysis. In particular, compared with primary data sets, generally fewer resources can be devoted by the suppliers of these data sets to ensure data quality and consistency.

Many of the criticisms regarding quality and consistency in secondary income distribution data put forward by Atkinson and Brandolini (2001) and later by Francois and Rojas-Romagosa (2005) are still valid. More generally, it also has been argued that survey estimates that build the basis for both primary and secondary data sets often only partially portray the income distribution (Pyatt, 2003). In addition, the fact that secondary data sets include indicators based on different concepts and definitions is often tackled by applying "dummy variable" adjustments. Atkinson and Brandolini (2001) conclude that such adjustments are not satisfactory because "differences in methodology may affect not only the level but also the trend of variables (so that it may not be sufficient to apply a fixed-effect correction in panel data estimation)" (p. 295). Rigorous sensitivity analyses are therefore required because data choices can impede both levels and trends in distributional indicators, which in turn can greatly affect the identification and interpretation of causal factors in an internationally comparative context. Primary users of the databases discussed above should not take the series collected at face value; they need to carefully

[16] Such a procedure can, however, occasionally result in dubious estimates, especially for earlier periods for which data sources are rare and less comparable.

examine the downloaded data. In turn, secondary users of the research based on one or the other of these databases ("meta-users") need to verify to what extent the researchers validated the data they used.

19.4. DEFINITION OF INEQUALITY MEASURES AND THEIR VARIABILITY

19.4.1 Definition of the Dependent Variable

This section describes how the dependent variable—household income inequality—is measured in the empirical work under review. It is important to note right from the outset that in an overwhelming majority of cases researchers do not have full discretion over which inequality measure they will analyze or include in their models. This is, in most of the cases, limited by the availability of the data, and it is especially so in the case of country-level comparisons of secondary data. The variable list of the large international secondary data sets (such as WIID, for example) hugely constrains the choice. The larger the data set in terms of country coverage, the more this is likely to be the case (because the possibility of having new, harmonized indicators diminishes with the size of the surveys). There are only a few measures usually available, of which the Gini coefficient is by far the most often used, followed by various decile shares (S80/S20 or S90/S10) and, sometimes, percentile ratios such as P90/P10 or ratios of some other percentile values.

None of the above-mentioned measures are overly sensitive to the tails of the income distribution, and therefore the analyses based on them may miss important changes within the distribution. This could partly be overcome by the use of more tail-sensitive measures such as D9/D5 ratios, generalized entropy-type measures of inequality (Theil, MLD), or Atkinson-class measures. However, it also became important to pay attention to polarization measures comparing the values of a comparison distribution to the values of a reference distribution (Alderson and Doran 2013; Handcock and Morris, 1999; Morris et al., 1994; Wolfson 1997). The share of population classified by cutpoints of the comparison distribution can show how it falls in similarly defined categories of the reference distribution, allowing us to compare relative positions of people at various parts of the distribution.[17]

Studies investigating developments of tail-sensitive overall inequality measures or polarization measures, however, remain rare in the literature, given the fact that these measures are, unlike Gini coefficients, much less available for international comparisons.[18] On the

[17] For some analyses inequality is measured by the relative welfare-to-material-to-income ratio of various social subgroups (elderly/children, higher educated/lower educated, gender, etc.).

[18] A recent attempt to construct a more tail-sensitive measure is the one suggested by Palma (2011). The Palma index compares the top decile share with the share of the bottom four deciles and is suggested to better reflect developments in the upper tail compared with the majority. However, its calculation requires the availability of decile shares (i.e., generally microdata), and because the top decile average—especially in small samples—is very vulnerable to accidental inclusion of outliers, care is warranted with the Palma index as well.

other hand, using the Gini and other middle sensitive measures does also have advantages, especially when sampling variability due to small sample sizes is an issue.

Further, in some studies, such as, for example, political science explanations, or in analyses of the effects of redistribution, it is not the actual value of the inequality measure such as the Gini coefficient (of net disposable incomes) in itself but the difference between the pre-tax and -transfer Gini on the one hand and the post-tax and -transfer Gini on the other that is used as the dependent variable. This is a measure of redistribution for many analytic papers (e.g., Bradly et al., 2003; Iversen and Soskice, 2006) and a proxy of how politics and policies affect inequalities.

The range of available inequality indicators also constrain the features of inequality that can be analyzed in international comparisons. If only inequality measures insensitive to the tails are available and analyzed, there is a risk that important changes in the income distribution are missed or noticed too late.

19.4.2 Variability of the Dependent Variable

Trends and patterns of inequalities in countries in the OECD area are analyzed in depth in Chapters 7–9 of this volume. Overviews of the developments of income inequality have been presented in a large number of studies; some of the recent core publications include OECD (2008, 2011),[19] Alderson and Doran (2013), Brandolini and Smeeding (2009), Ward et al. (2009), Tóth (2014), Ferreira and Ravallion (2009), Salverda et al. (2014), and Nolan et al. (2014).

One of the most fundamental questions of comparisons of inequality is the variability of the measures used to characterize inequality in society, both across countries and over time as well. The large and rapidly growing income distribution literature (Atkinson and Bourguignon, 2000; Salverda et al., 2009) presents various narratives about the development of income inequality. The major narrative dominating the literature is proposed by the landmark studies of the OECD (2008, 2011) and by various papers based on the data collections of the LIS. According to this, within-country inequalities have increased in a majority of OECD countries since the 1980s, and at least until the breakout of the Great Recession (OECD, 2008, 2011, 2013a; see also Atkinson, Rainwater and Smeeding, 1995; Gottschalk and Smeeding, 2000; Brandolini and Smeeding, 2009; Chapter 8 of this volume). As the most recent OECD (2011) study stresses, in a large majority of OECD countries the income of the richest 10% of households has grown faster than that of the poorest 10%. The Gini coefficient increased on average from 0.286 in the mid-1980s to 0.316 in the late 2000s. Of the 22 countries for which a long time series is available, 17 have witnessed increasing inequality. For seven of these the Gini coefficient increased by more than four points over the period. In only five of these countries

[19] For a summary, see Förster (2013).

did inequality not increase or even decline. This is a narrative proposing inequality trends, which are dominant in the era of the "great U-turn" of inequality developments.

After an analysis of the GINI Inequality and Poverty Database, Tóth (2014) concludes that over the past three decades, inequality has indeed increased on average across the countries included in the analysis (25 EU countries, to which the United States, Canada, Korea, Japan and Australia are added); the whole range of Gini coefficients were at a higher level at the end of the period (from a minimum/maximum level of 0.20/0.33 to 0.23/0.37). The above work also stresses that the growth in inequality was far from uniform. In some countries (mostly in continental European welfare states such as Austria, Belgium, France), the level of inequality remained largely unchanged or fluctuated around the same level, whereas in others it increased substantially. The latter trend was experienced by some European transition countries (Bulgaria, Estonia, Lithuania, Latvia, Romania and Hungary) and to a lesser but still a considerable extent by the Nordic countries, most notably Sweden and Finland. It also was found that the pattern of inequality change may sometimes show declines for shorter or for longer periods. Such spells of decline were observed in Estonia, Bulgaria and Hungary, for example, sometimes after sharp increases.

Finally, over time it seems possible indeed that countries shift between inequality regimes (Tóth, 2014). After decades of a gradual but incessant increase of inequality, some of the Nordic countries, for example, while still being part of the group of low-inequality countries, no longer are at the lowest end of the inequality "league table." The United Kingdom moved from being a middle-level inequality country in the 1970s to the group of high-level inequality countries by 1990. Also, some of the transition countries such as the Baltic countries, Romania or Bulgaria witnessed very large changes that have put their inequality levels in a different range (see also Tóth and Medgyesi, 2011). Chapter 8 of this volume provides a more detailed account of post-1970 trends in within-country inequality in OECD and a range of middle-income countries.

19.4.3 Reliability of the Dependent Variable

Population surveys from which data on inequality are computed cover only a sample of the population. Originating basically from this fact, there is always a sampling variance of the statistic chosen to describe features of the distribution. The variability of the sample estimate about its expected value in hypothetical repetitions of the sample (the sampling variance) may be due to sampling and nonsampling errors. Most surveys are based on complex sample design (allowing, for example, a stratification of base populations to draw the sample, of a clustering of cases, of differential techniques providing equal probability of getting into the sample, etc.) Nonsampling errors (of coverage, wording, nonresponse, imputation, weighting, etc.) add to the uncertainty of the selected statistics.

All inequality measures (Gini figures, P90/P10 ratios, etc.) used in international comparisons are estimates from samples that are, in most cases of different designs, based on partially (or not at all) harmonized surveys. In addition, inequality indices are not like simple ratios from samples; for most of them the calculation is based on complicated formulae, leading to nonlinearities of the indexes. It is therefore very important to understand to what extent secondary uses (i.e., multivariate and multicountry analyses of drivers of inequality) can account for such uncertainties.

Inference for inequality and poverty measures calculated from properly documented microdata can be tested by "direct" or formula-based (asymptotic) methods and by experimental methods (based on resampling techniques such as bootstrapping or Jacknife, for example) (see Kovacevic and Binder, 1997; Biewen and Jenkins, 2006; Osier et al., 2013; and others). Both types of methods are used in various research contexts, but none of the results are frequently reported in official statistics and in secondary datasets. While it is shown that the way inference is calculated is important—Davidson and Flachaire, 2007, for example, found that in the case of complex sample design, bootstrapping may lead to not accurate estimates of inference, even for very large samples—sticking to point estimates only is clearly problematic, in part because it creates false images of certainty in inequality statistics and in part because it misguides interpretations of intertemporal change and cross-country differentials. While the degree of accuracy that may be worth pursuing is open to discussion (as Osier et al., 2013 stress, there is need to address a trade-off between statistical accuracy and operational efficiency when choosing estimation methods for standard errors), overlooking the issue is clearly the worst option.

To properly estimate sampling variance, sample design, weighting procedures, imputation practices and the actual computation formula of the statistic is to be taken into account. The effects of these factors are tested in various papers. As Goedemé (2013) and Biewen and Jenkins (2006) stress, ignoring the effect of clustering of individuals in households for poverty indexes (that are derived from incomes measured at the household level but analyzed at the individual level) may lead to a serious underestimation of standard errors for the analyzed poverty measures. Taking clustering into account leads to fairly good proxy of "true" estimations to settings when sample design variables are not missing. Little is known on similar tests for inequality measures.

Van Kerm and Pi Alperin (2013) tested how their measures of inequality reacted to the presence or elimination of extreme values from the surveys they analyzed, and they found their measures were arbitrarily large when they left outliers in their sample. However, other measures such as poverty rates remained more robust for the presence or elimination of extreme values (Van Kerm, 2007).

An essential requirement for computation of variance estimates for inequality measures is that microdata be available for analysis. Most secondary data sets lack any indication of not only the standard error estimates but also essential properties of the samples they have been drawn from. This makes it especially difficult for comparative studies using secondary data sets to assess reliability of their findings.

Further, the Gini coefficient, by construction, is a variable with a relatively small range. Even if inequality may change significantly in the long run, when shorter periods are taken into account and when many data points within the longer period are considered, the adjacent Ginis (in time or across countries) may not (in statistical terms) be significantly different from each other. Therefore, if these values are put into a variable on the left-hand side of a regression, there is a serious risk that a large "noise" enters the estimates.[20]

Also, when using secondary datasets, where there are no microdata at hand the researchers have to apply some rule of thumb to decide what can be considered a "real" change over time. There is no agreement in the literature, however, about how over-time changes or cross-country differences of Gini coefficients (normally arrived at from heterogeneous sample designs and greatly varying samples) could be defined as significant in statistical terms. Bootstrap (or, better, linearization) estimates of confidence intervals of Gini would suggest roughly ±1 Gini point differences in EU-SILC samples to be registered as "significant," but little is known on how this could be applied to changes over time given the lack of information in necessary detail about sample designs.

Atkinson (2008) proposed a simple metric of changes in the case of considering changes in percentiles (relative to the median) over a period of decades. He requires a 5% change to be "registered," a 10% change to be qualified as "significant," and a 20% change to be qualified as "large." The bottom decile falling from 50% to at least 47.5% of the median thus would "register" as a change, be considered "significant" if falling below 45%, and being considered "large" if falling below 40%.

Breaks in series pose a serious challenge for cross-country comparisons as well as for intertemporal tracking of inequality, as already noted (Atkinson and Brandolini, 2001). A break in a series may provide an obvious basis for suspicion if accompanied by a sudden change in the level of inequality that subsequently does not continue in the same direction. However, in other cases one must rely on expert judgements as to whether such breaks have in fact masked an underlying change in inequality.

A way of constructing long-term data series of inequality is to link subsequent data series stemming from different data sources or definitions together with use of information on overlaps of these series (Atkinson and Morelli, 2014; Förster and Mira d'Ercole, 2012), a method often called "data splicing ."[21]

A proper definition of *inequality change* in empirical studies (in addition to knowledge of sample sizes and sample designs) also has to be based on careful examination of annual increments of the inequality measure at hand, on the length of the data spells, and on

[20] In general, the articles reviewed do not publish confidence intervals for the inequality measures.

[21] The OECD Income Distribution Database (IDD) described earlier also applies data splicing (when needed) (see OECD, 2013b).

many other "accidental" factors. As Tóth (2014) stresses, a year-to-year difference up to a magnitude of 1 Gini point can be considered as no change, especially if variation in subsequent years go in different directions. However, consistent year-to-year changes, even if small ones (say, half a point) from 1 year to another, may accumulate into a five-point change or more in the Gini over 10 years, which is a substantial change indeed. Such longer-run consistency of increments over time may also change the interpretation of short-term comparisons. Consider long-term fluctuation of the Belgian or the Irish Gini series (resulting in longer periods of "no change" in inequalities) and compare those with the very small but consistent year-to-year increments of Ginis in Sweden or Finland, and it becomes clear how important it is to pay attention to even small and insignificant Gini changes (Tóth, 2014).

Nevertheless, when the Gini index is used as left-hand variable in regressions, spell contexts (as defined above) cannot always be taken into account, and the actual interpretation of the parameter estimates depends heavily on statistical inference. Careful and balanced evaluation: this is the main lesson that can be drawn and the only suggestion that can be given at this stage.

19.5. DRIVERS OF INEQUALITY: MAIN EXPLANATIONS

This section sets out the main arguments of inequality drivers in OECD countries put forward in cross-country studies and reports the results from recent empirical work supporting or not supporting these arguments. We focus our review of the literature including studies undertaken in the past 10–15 years, with no pretention of exhaustiveness. In particular, this review updates Atkinson and Brandolini (2009) and extends the literature review by Chen et al. (2013a).

The section introduces the main factors put forward to explain international differences in levels and trends of income inequality. The discussion is structured along six main headings: structural macroeconomic sectorial changes; globalization and technical change; changes in institutions and regulations; political processes; redistribution via taxes and transfers; and structural societal changes. Annex Table A19.1 gives an account of the wealth of findings for a subset of 48 selected studies that are considered to be the most pertinent ones undertaken in the past 10–15 years. The selection criteria relate to coverage (i.e., the studies should include a critical mass of countries and should focus on the joint OECD and EU areas); multivariate explanations (i.e., monocausal studies were excluded); and timeliness (i.e., preference was given to more recent studies not yet included in literature surveys available elsewhere).

When talking about "main drivers," it is useful first to make a distinction between direct, or proximate, drivers and indirect, or underlying, factors resp. causes behind changes in income distribution (see Cornia, 2012 for the same distinction). Direct drivers can be gauged, for instance, by decomposing summary income inequality measures by income components or by calculating the first-order effect of changing household structures on income distribution, for example, by using shift-share analyses. A variety of such direct factors for growing inequality in OECD countries has been identified by the OECD (2008). While usually analyzed in isolation, such identification of factors—especially if as exhaustive as possible—provides a useful checklist of "hints" (Cornia, 2012) at indirect factors or causes that lie behind inequality changes. In the following subsections, we classify the main underlying factors into six overall headings, following the presentation in Figure 19.1.

The subsections below resume the arguments put forward in the literature and report the results from empirical analyses. The main "culprits" tested in the literature have been subsumed under the different subheadings enumerated above, each observing single sets of drivers of inequality and inequality changes, thus defining more monocausal explanations of inequality. Of course, none of the studies reviewed is monocausal in nature, and all test the significance and relative importance of several drivers, but the point of departure is often related to one particular area, for example, the impact of globalization versus technology or versus institutions.

Our review focuses on OECD and EU countries. The country coverage in some studies is limited to only a subset of OECD countries, whereas many other studies include a larger sample of countries, including notably middle-income and developing countries. Given the focus of this chapter, we review below results pertaining to the OECD area, also when obtained from the second strand of studies insofar as results for OECD countries are reported separately.

Though our preferred explanatory variable is dispersion of household disposable income, we also report findings that explain changes in the distribution of earnings. While the use of one or the other of these two income concepts may alter the findings (net income estimates also are affected by household structure and tax/transfer changes), and definitions within these two aggregates differ (full-time wages or annual earnings; gross or net incomes), a number of studies refer exclusively to the effect on earnings, especially those looking at the causal role of trade and technology. Findings referring to income and earnings are presented separately below.

19.5.1 Structural Macroeconomic Sectoral Changes

For a long time, the quest to identify driving factors of inequality looked primarily at the association between economic development and inequality and was focused on testing the hypothesis that Kuznets (1955) put forward. According to this hypothesis, inequality follows an inverted *U*-shaped relationship with increased development. This is linked to a sectoral move from a "traditional" sector (agriculture) to a "modern" sector (industry). Insofar as the traditional sector is less productive, it will provide lower wages than the modern sector (sector dualism); it also is expected that the traditional sector has lower inequality within it (sector bias). Consequently, it is expected that development first increases and subsequently decreases inequality.

Usually, economic development is proxied by real income or GDP per capita (y). To capture the parabolic shape of the relationship, the quadratic form of y is added. Following Hellier and Lambrecht (2012), in the frame of a panel of country studies, the relationship can be written as:

$$INEQ_{i,t} = \alpha + \beta_1 y_{i,t} + \beta_2 {y_{i,t}}^2 + AX_{i,t} + \varepsilon_{i,t} \tag{19.3}$$

where i and t are country and time, y is per-capita real income (or GDP) and $X_{i,t} = \{x_{j,i,t}\}$ a vector of variables j that affect the inequality measure $INEQ$. These variables seek to control for shocks as well as institutional and regulatory differences across countries. Equation (19.1) is a specific variant of the general regression equation GIRE described earlier in Section 19.2.1. The Kuznets hypothesis then is confirmed if the estimated values β_1 and β_2 are such that $\beta_1 > 0$ and $\beta_2 < 0$. The turning point, where inequality attains its highest value and begins to decrease, can then be estimated to correspond to the period $\boldsymbol{\Omega}$, such that $y_{\boldsymbol{\Omega}} = y_0 - \beta_1 / 2\beta_2$ (for a start of the estimation at time $t = 0$ with the income per capital y_0).[22]

Evidence from studies of the inequality/development relationship remains broadly inconclusive. Around half of the studies reviewed by Atkinson and Brandolini (2009) estimate such relationship, with or without other controls. Some of these studies support the Kuznets hypothesis but others reject it. Hellier and Lambrecht (2012) undertake a review of studies testing the Kuznets hypothesis. Studies based on cross-sectional analysis of countries in their majority tended to support the Kuzents hypothesis (although some clearly reject it), whereas the evidence from panel data estimations is more mixed. In a study of the EU member states between 2000 and 2005, Medgyesi and Tóth (2009) suggest absence of a clear relationship between the economic growth rate and inequality within EU member states in the first half of the 2000s. Bourguignon (2005) concludes that, overall, the analyses of the available data at hand "do not suggest any strong and systematic relationship between inequality and the level of development of an economy" (p. 1733).

Empirically, the past 20–30 years were characterized by a considerable increase in earnings and income inequality in a large majority of OECD countries (OECD 2008, 2011), a development that is sometimes called "the great U-turn" (but see Section 19.4.2 on variability of inequality measures). Even if one considers the

[22] In discussing the appropriate specification of the Kuznets relationship, Anand and Kanbur (1993) derive functional forms of and conditions for the turning point for six different inequality indexes. They show that under the Kuznets assumptions, different indices of the Lorenz class increase at the start of the development process, but the behaviour at the end of the process—and the existence of a turning point—is ambiguous. Importantly, each index is shown to have its own functional form and turning point condition.

inequality/development relationship to be accurately described as an inverted *U*-shaped curve, this picture needs to be amended and replaced by an *N*-shaped (Alderson and Nielsen, 2002) or tilde-shaped (Hellier and Lambrecht, 2012) curve.

Alderson and Nielsen (2002) test the Kuznets hypothesis by applying a measure of sector dualism (shift of employment out of agriculture) for 16 OECD countries for the period 1967–1992. They find that sector dualism has no significant effect on income inequality unless none of the globalization variables are controlled for. At the same time, sector bias (measured as the share of the labor force in agriculture) has a strong and positive effect. The latter surprising positive sign is explained by Alderson and Nielsen by the fact that dualism in agriculture has become less relevant for OECD countries for overall inequality, and its meaning now is more likely to be a measure of agrarian traditionalism than a component of the dualism model.

The "great U-turn" may then better be explained by other phenomena such as globalization or institutional change (see the next section). Still, issues of sector dualism and sector bias can be expected to play an important role when analyzed in terms of a sectoral change from a postindustrialized to a knowledge society. Nollmann (2006) and Rohrbach (2009) propose a model similar to that of Alderson and Nielsen (2002) but focus on sector dualism in terms of the wage differential between the knowledge sector and the remainder of the economy and on sector bias in terms of employment shares in the knowledge sector. For a panel of 19 OECD countries for 1970–2000, Rohrbach finds support for the sector bias hypothesis but no support for sector dualism. Moreover, and in contrast to Alderson and Nielsen (2002), Rohrbach (2009) finds no significant effect of globalization (in terms of trade openness), concluding that factor effects remain central determinants for understanding inequality. This traces back to the original argument by Kuznets that through the segmentation of factor markets sectorial changes can be important drivers of inequality changes. However, while there is some segmentation of the labor market in OECD countries, it does not appear across large sectors of activity. The high-tech/low-tech distinction seems more important but less easy to implement analytically.

19.5.2 Globalization and Technical Change

Since the 1990s, economic globalization has been intensively analyzed as one of the main potential drivers of increased earnings and income inequality in the OECD area. "Globalization" is, however, a multifaceted phenomenon and cannot be reduced to a single variable.[23] There are different aspects of it and they are likely to affect trends in earnings and income inequalities in different ways and in possibly opposing directions:

[23] Note that the discussion here and later refers to the "new age" of globalization (or Globalization II). It has been suggested that the distributive effects of the earlier Globalization I during the late nineteenth century up to World War II have been very different (Milanovic 2012).

- trade integration (goods and services mobility)
- financial integration (capital mobility)
- production relocation (firm mobility)
- technology transfers (information mobility)
- political aspects of globalization

The following subsections consider these aspects in turn.[24]

19.5.2.1 Trade

Increased trade integration is often taken as a main sign and sometimes as the sole proxy for the degree of economic globalization. The share of world trade in world GDP has grown from about one third to over half in the past 30 years (IMF, 2007). In most OECD countries, the extent of trade integration has doubled or tripled during this period, and the increase was especially stark during the 1990s (OECD, 2011).[25]

The standard reading of traditional international trade theory is that increased trade integration is associated with higher relative wages of skilled workers in advanced countries, thus contributing to increased inequality in those countries and higher relative wages of unskilled workers in developing countries with an associated decrease in inequality (for a discussion of the relationship between skill differentials and globalization, see, for instance, Krugman, 1995, 2000 and Kremer and Maskin, 2003). This is based on predictions of the Heckscher–Ohlin (HO) model, or variants of it. This model expects that countries export goods that use intensively the factor with which they are most abundantly endowed and import those that intensively use their scarce factors. Advanced countries with abundant highly skilled labor will therefore import products from countries with lower endowments of skills and export products made by skilled workers. Combined with the Stolper–Samuelson theorem, which predicts that trade increases the real returns to relatively abundant factors, increased trade integration should then reduce the demand for less-skilled workers and increase the demand for skilled workers in advanced countries and the inverse in developing countries Heckscher-Ohlin-Samuelson (HOS) model. Second, less-skilled workers are predicted to migrate to advanced countries. Third, capital would flow from advanced countries with large capital-to-labor ratios to developing countries with small capital-to-labor ratios. All three processes are predicted to lead to increased inequality in advanced countries and to decreased inequality in developing countries.

However, most studies found it difficult to reconcile the empirical evidence on earnings and income inequality trends with the traditional HOS model, which typically does not capture technology diffusion. A number of cross-country studies find trade globalization

[24] There are additional features of globalization that may have indirect and direct effects on the distribution of income, such as cultural aspects of globalization or migration, which are, however, beyond the scope of the detailed discussion in this chapter. The issue of migration is discussed partially (as a trend having composition effects on societies) in Section 19.5.6.

[25] Note, however, that the increase in the GDP share of trade would be much lower if trade was measured in terms of value added.

to have increased income inequalities in high-wage and low-wage countries alike, which is at odds with traditional trade theory (for a review, see Milanovic and Squire, 2007). Furthermore, all sectors tended to become more skill intensive (as already reported by Krugman, 1995). Chusseau et al. (2008) relate this to the fact that trade between advanced and developing countries still accounts for a lower share than trade between advanced countries, thereby playing a lesser role in the shift of factor demand (Chusseau et al., 2008).

Some of the shortcomings of the traditional HOS model have been put forward by, for instance, Davis and Mishra (2007). The particular assumption of growing capital flows from developed to developing countries and their equalizing impact (in developing countries) has been challenged, notably on the grounds of capital market imperfections (Lucas, 1990; Alfaro et al., 2008). During the past 15–20 years, new approaches in trade models have been developed to overcome analytical shortcomings of the HOS model in several areas. The first one is to take account of heterogeneity of firms within industries in both developed and developing countries based on the development of dynamic industry models, as in the work of Melitz (2003). The coexistence of more productive firms that are expanding and entering the export market and contracting less productive firms within the same industry has an effect on how trade influences the wage and income distribution (Pavcnik, 2011). Exporting firms can employ more productive workers and offer higher wages, with a possible sizeable effect on increased wage inequality within sectors.

This calls into question the assumption of competitive labor markets underlying the HOS model, which expects an equalizing wage distribution in developing countries through higher unskilled wages. Newer trade theories therefore accounted for labor market imperfections by including efficiency wage models or models of fair wages in their framework (e.g., Verhoogen, 2008; Egger and Kreickemeier, 2009, 2010). In a next and complementary step, attempts were made to relate the exporting firms' wage premium to search frictions as a source of labor market imperfection, introducing search and matching models (Helpman et al., 2009). In both streams of work, trade liberalization can be consistent with increasing residual wage inequality, that is, inequality between workers with the same skills and other characteristics.

Empirically, however, both these channels, which are related to the recognition of heterogeneity of firms, can only be observed and analyzed at the micro level, going beyond models based on "representative firms." A number of studies reporting results for particular countries, mainly Latin American countries and Indonesia, were published in the later 2000s. Most of these studies (reviewed by Pavcnik, 2011) suggest that increased export market access was associated with greater wage inequality in a given country. But there are no cross-country studies available so far.

There are channels other than the HOS model through which trade can affect income inequality. One is increased competition, which tends to reduce the relative prices of consumption goods and can also diminish the monopoly position enjoyed by the upper class—both processes would reduce income inequality (Birdsall, 1998). A more indirect argument refers to the second-order effects of decreases in the relative wages of unskilled workers; this may lead to incentives for workers to up-skill and for employers to hire more unskilled labor, leading to lower inequality (Blanchard and Giavazzi, 2003). There are also other theories and models that predict that inequality would decrease in both advanced and developing countries, namely through the effect of specialization; such division of labor could generate increasing returns to scale, whereby labor has a higher marginal productivity (Francois and Nelson, 2003).

In the following, the empirical results of selected pooled cross-country studies are summarized, distinguishing effects of trade globalization on wage dispersion on the one hand and on income inequality on the other. When discussing the effect on wage dispersion, the notions of "wage differential" and "wage distribution" need to be distinguished. The models described above (in particular the HOS theory) yield predictions about the wage differential (i.e., on wage ratios between various skill or occupation groups), but the effect on the distribution of wages also depends on quantities (i.e., the number of people earning these wages). If quantities are fixed (as assumed in a static trade theory), one can read the distribution of wages directly from the wage differential. But if people migrate and change across sectors, one cannot predict distributional effects directly from changing wage differentials. Most of the empirical studies reviewed below test the potential effect of trade integration on wage distribution.

19.5.2.1.1 Wage Dispersion Effects

For a set of 23 OECD countries 1980–2008, OECD (2011) suggests that trade integration[26] has no significant effect on trends in wage dispersion at the aggregate level within countries once the effects of technological change and institutions are controlled for. This result holds for both top and bottom sensitive indicators of earnings (interdecile ratios) and when imports and exports are examined separately. An insignificant distributive effect of trade integration is also estimated for the overall earnings distribution among the entire working-age population (i.e., including the unemployed), insofar as trade had neither a significant positive or negative effect on employment.

On the other hand, Cassette et al. (2012) suggest a positive relationship between trade and wage dispersion for a subsample of 10 OECD countries between 1980 and 2005, which, however, differs between goods and services as well as in short- and long-run estimates. In the short run, wage dispersion is widened by increased trade in goods, whereas trade in services has no effect. That differs from long-run effects, where trade in services increases inequality, in particular at the top of the earnings distribution (i.e., between top and median earnings).

For OECD countries, a subaggregate of total trade may be a more pertinent indicator, namely the share of imports from low-income developing countries (LDCs). However, Rueda and Pontusson (2000) suggest that its increasing share had no effect on wage

[26] Trade integration is measured as trade exposure, that is, a weighted average of import penetration and export intensity.

dispersion, at least for the period up to 1995. Similarly, Mahler (2004) shows that, for a subset of 14 OECD countries for the period 1980–2000, imports from LDCs had no significant distributive effects on either earnings or disposable incomes. For the more recent period up to 2008, OECD (2011) reports similar findings, although with nuances: overall the effect of LDC imports is distribution neutral, but considering the institutional context, such imports tend to compress the wage dispersion in countries with stronger employment protection legislation (EPL) but widen it in countries with weaker EPL. For Golden and Wallerstein (2011), however, trade with LDCs is one of the key drivers of increased wage dispersion within 16 OECD countries during the 1990s.[27] Their results distinguish the period of the 1990s from the decade of the 1980s, when trade played no role but institutions did (see Section 19.5.3). Among those finding a moderate disequalizing role of imports from LDCs are Alderson and Nielsen (2002), although their results refer to income rather than earnings inequality.[28]

19.5.2.1.2 Income Distribution Effects

Few studies estimate the effect of trade openness for the group of OECD countries on the distribution of income directly. For the subgroup of advanced countries analyzed by the IMF (2007), economic globalization overall (trade and financial globalization taken together) contributed to increasing income inequality, but this was entirely because of foreign direct investment (FDI) trends, which more than outweighed the equalizing effects of trade: both exports and, in particular, imports from LDCs (but not trends in tariffs) were associated with decreasing income inequality in advanced countries. Similarly, for 24 OECD countries for the period 1997–2007, Faustino and Vali (2012) found that trade liberalization decreases income inequality, making use of both static and dynamic regression estimates. In a study of 16 OECD countries, the ILO (2008) included tariff liberalization as only a proxy for trade openness, finding no significance for an effect on income inequality.

19.5.2.2 Trade Openness and Inequality in an Enlarged Country Sample

There are somewhat more findings attributing distributive effects to increased trade integration when the country sample is enlarged from the group of OECD countries.[29]

[27] Their results suggest a one percentage point increase in trade, with LDCs being associated with a one percentage point increase in wage inequality.

[28] Results suggest that increasing LDC import penetration by 1 standard deviation increases the Gini coefficient of income inequality by 0.6 points.

[29] When analysis is restricted to the OECD area, a group of relatively homogenous economies in terms of their development status, it is reasonable to disregard differences in national income levels when assessing the contributions of factors such as trade globalization on the income distribution. Enlarging the country sample, however, needs to take into account that trade and other globalization variables may have different effects on inequality depending on a country's level of development. That is what is at least predicted by the traditional HOS theorem or variations of it. Estimating the effects of globalization on income distribution in both richer and poorer countries together therefore requires analysis of the interaction with GDP/capita and economic growth.

Evidence is mixed, and for a full sample of 129 countries for three points in time in the 1980s and 1990s, Milanovic (2005) suggests that as national income increases, the inequality effects of globalization reverse, enhancing inequality at poorer income levels but dampening inequality at higher levels.[30] This runs counter to the hypotheses of the classical HOS model.

Milanovic and Squire (2007) investigated the effect of trade (measured with the unweighted average tariff rate) on interoccupational and interindustry wage differentials for the period between 1980 and 1999. For both indicators, a decrease in tariff rates tended to have a positive association with wage dispersion in poorer countries but a negative association in richer ones. Institutions (union density and coverage) do not play a role in interoccupational wage disparity but reinforce the disequalizing effect on interindustry wage differentials.

For a panel of 51 countries, Bertola (2008) found that trade openness is positively associated with inequality of both gross income and disposable income (for a smaller set of countries) and that government expenditure is less redistributive in countries with a higher degree of trade openness. Spilimbergo et al. (1999) suggested that the effects of trade openness on inequality depend on factor endowments, increasing income inequality in skill-abundant countries but reducing it in capital-abundant countries. Based on newer data and a larger country sample, Gourdon et al. (2008) nuanced this finding. Measured as a lagged ratio of tariff revenues to imports, they found that trade openness is associated with increases in income inequality in both high skill-abundant and capital-abundant countries. By contrast, IMF (2007) suggests that the role of trade globalization in the last two decades of the twentieth century was insignificant overall, but some elements actually contributed to decreasing income inequality, in particular lower tariffs and higher agricultural exports.

For the specific country group of Latin American countries, Cornia (2012) found, perhaps contrary to expectations, that the gains in terms of trade realized during the 1990s and 2000s contributed significantly, albeit modestly, to the recent decline in income inequality. This is explained by relaxed external constraints on growth and consequently increased incomes, employment, and revenue collection.[31]

19.5.2.3 Financial Openness

There are mechanisms other than trade through which economic globalization can accelerate earnings and income inequality. One such mechanism is cross-border movement of capital, a factor that is overlooked in the basic trade model, which assumes that labor and capital are mobile within a country but not internationally. Factors such as deregulation,

[30] Milanovic (2005) identifies the "turning point" as around US$8.000 per capita in 1985 PPPs.

[31] However, the reversal of the skill premium as well as a shift towards more progressive labor and fiscal policies are identified as the main factors for the decrease in inequality (Cornia, 2012).

privatization and advances in technology all contributed to the rapid growth of capital movement, in particular FDI, over the past decades. If the utilization of capital as well as embodied technology requires the use of skilled workers, and capital and skilled labor are complementary, the increase in inward capital will increase demand for skilled workers (Acemoglu, 2002).

Much like HOS models of trade, models of FDI usually predict different effects in advanced and developing countries. If FDI flows are directed to countries with relative abundance of low-skilled labor, this should a priori increase the demand for the abundant factor and hence have an equalizing effect in developing but a disequalizing effect in developed countries. However, less skill-intensive outward FDI from advanced countries can appear as relatively high skill-intensive inward FDI in developing countries. In that case, even when the transferred technology is "neutral," an increase in FDI from advanced to developed countries can increase the demand for skilled labor and contribute to increasing inequality in both advanced and developing countries (Feenstra and Hanson, 2003; Lee and Vivarelli, 2006). Further, there may be indirect disequalizing effects, even if FDI is mainly attracted by low skill-intensive countries and sectors; to attract FDI, countries may relax regulations in the field of employment protection or fiscal parameters, which otherwise would have an equalizing effect (Cornia, 2005).

Endogenous growth models such as those proposed by Aghion and Howitt (1998) or Aghion et al. (1999) assume two stages of development and inequality when new technologies are introduced: in the transition phase skilled labor demand and hence wage inequality increase before decreasing in a second stage. Such models can be adapted in terms of effects of FDI on the availability of new technologies. Figini and Görg (2006), for instance, view FDI as a vehicle for introducing new technologies. They expect that in a first step more FDI will lead to increased inequality between skilled and unskilled workers, with a reversed trend in the second step as domestic firms follow up imitating advance technologies.

19.5.2.3.1 Wage Dispersion Effects

Figini and Görg (2006) wrote one of two articles in our review that use FDI as the main explanatory factor for distributional changes. Their model specifies only the inward component of FDI. For the subsample of 22 OECD countries, they found that higher inward FDI is significantly (at the 5% level) related to lower earnings inequality in the manufacturing sector for the period 1980–2002. Further, this effect seems to be linear. This is in contrast to the results for non-OECD countries, where the inward FDI has a positive though nonlinear association with earnings inequality.

Similar findings are also suggested in the results of OECD (2011) for 23 OECD countries between 1980 and 2008. Although overall FDI turns out to be insignificant, inward FDI has a significant equalizing effect on wage distribution and outward FDI has a disequalizing effect, although the latter effect is rather modest (see the next section). Inward

FDI, however, seems to be correlated with trends in trade integration. Other indicators of financial openness were reported to be insignificant in this study; this concerns cross-border assets and liabilities, foreign portfolio investment, and a de jure measure of FDI restrictiveness, which was the preferred measure of financial openness in this study.[32]

Among more country-specific studies, Taylor and Driffield (2005) found that inward FDI flow can explain, on average, 11% of the increase in wage inequality in United Kingdom between 1983 and 1992. Bruno et al. (2004) examined the effects of inward FDI on relative skilled labor demand and wage differentials in manufacturing in the Czech Republic, Hungary and Poland for the years 1993–2000. They found that FDI did not contribute to increasing wage dispersion in the three countries, although it did contribute to increasing the skill premium in the Czech Republic and in Hungary (but not in Poland). Hijzen et al. (2013) analyzed microeconomic (firm-level) data for three developed and two emerging economies and found that wage premium effects following foreign ownership are larger in developing countries, that the largest effect on wages comes from workers who move from domestic to foreign firms and that employment growth after foreign takeover is concentrated in high-skill jobs.

19.5.2.3.2 Income Distribution Effects

Most studies reviewed found only modest or no significant effects of overall FDI in OECD countries, but there are more significant results when inward and outward FDI are analyzed separately. Using time series data for the period 1960–1996, Reuveny and Li (2003) showed that inward FDI flow for 69 countries is significantly and positively associated with income inequality for both OECD and less developed countries, which were sampled separately. The IMF (2007) reached the same conclusion: for the subsample of advanced countries in the study of trends over 1980–2003, they identified both inward and, in particular, outward FDI as the elements of globalization that most increased income inequality, slightly more than outweighing the equalizing effect of increased trade. For a more recent period, 1997–2007, increased inward FDI was also found to be significantly positively related to income inequality for a sample of 24 OECD countries by Faustino and Vali (2012).[33] This seems to back up the observation that FDI occurs in more skill- and technology-intensive sectors.

The opposite was found by Çelik and Basdas (2010). Their article is the second of the two studies in our review that uses FDI as the main explanatory factor for distributional changes. For a subsample of five developed countries, their analysis suggests that both FDI inflows and FDI outflows are associated with *decreased* income inequality for the period of

[32] This is because de facto volume-based measures of financial openness such as FDI or foreign portfolio investment are often endogenously determined by other factors included in the framework, for example, technology or trade, as has been shown above.

[33] The effect of FDI, however, becomes insignificant when the authors control for potential endogeneity by applying generalized methods of moments estimators.

the mid-1990s to mid-2000s. The working hypothesis is that this is attributable to greater redistribution permitted by higher tax revenues from increased employment in the case of FDI inflows and changes in the economic structure with low-skilled labor being pushed to up-skill in the case of FDI outflows. The small number of observations (5 countries for 11 time observations), however, casts some doubts on the robustness of the results.

On the other hand, the ILO (2008) estimates that the inward FDI share in GDP had no effect on income inequality in a sample of 16 OECD countries for the period 1978–2002, as long as the analysis controls for technology (information and communications technology [ICT] share)—otherwise FDI comes out as a significant predictor, suggesting that FDI could act as a proxy for that omitted factor and actually lead to greater demand for skilled labor.

Somewhat more clear-cut results were found for the region of Latin America. Cornia (2012) examined a subsample of 19 Latin American countries for the period from 1990 to 2009. Given the boom in capital inflow, Cornia expects deteriorating effects on income inequality via an appreciation of the real exchange rate and a dampened growth in the labor intensive noncommodity traded sector. Indeed, the FDI stock had a significant and strongly disequalizing effect in all specifications, and the effect is most pronounced among the group of Andean countries (where FDI is particularly important in the mining sector). That said, in this analysis FDI—such as other external economic and demographic variables considered—had a more limited average effect on income inequality than the policy variables.

A more disequalizing effect of FDI also often is found in studies with the broadest possible country coverage. Broadening the analysis to 42 advanced and developing countries, the ILO (2008) found inward FDI to be the only variable among eight economic controls to be robustly positively associated with increased income inequality. This positive association was confirmed by the IMF (2007) for 51 countries, although technology played an even stronger role in the latter study. Higher inward FDI benefits solely the top quintile, whereas income effects for the three bottom quintiles are significantly negative. For a panel for 111 countries from 1970 to 2000, Te Velde and Xenogiani (2007) showed that FDI positively affects skill formation not only within countries but also across countries, especially in countries that are relatively well endowed with skills to start with. On the other hand, in his analysis of 129 countries for three benchmark years (late 1980s, early 1990s, late 1990s), Milanovic (2005) found that FDI has no effect on the income distribution, whether alone or when interacting with income. However, results from analyses that pool developed and developing countries are difficult to interpret because this blurs the channels through which financial openness affects the distribution of incomes, especially when inward and outward FDI are netted out.

19.5.2.4 Outsourcing

Most of the evidence that relates increasing earnings or income inequality on increased trade openness focuses on trade in *final* goods. As shown earlier, a larger part of the literature suggests that trade, measured in these terms, has not been the major driving factor (if at all) of increased inequalities in the OECD area. Such findings, however, neglect that the production of goods itself has become globalized, and outsourcing in terms of increasing trade in *intermediate* products may play a decisive role. It has been estimated that the potential of off-shoring of tasks concerns between 20% and 30% of all jobs in a number of OECD countries, including medium- and high-skilled jobs; however, tradability is determined not only by the technical feasibility of unbundling and digitization but also by transaction costs and the economies of scope of keeping tasks together (Lanz et al., 2011).

Among the first to put forward the outsourcing hypothesis, Feenstra and Hanson (1996) suggested that the rapid development of international production sharing[34] (from home companies to their foreign affiliates) may distort the wage distribution in home countries by moving some of domestic non-skill-intensive activities abroad. Such a move concerns potentially all firms (not only traded industries) as long as business owners find the fragmentation of production more cost-effective. Firms in advanced countries may "outsource" particular stages of production to less developed countries; these stages seem less skill-intensive in the advanced country but relatively skill-intensive in the receiving country. As a result, trade—the outsourcing aspect of it—may reduce the relative demand for unskilled workers and increase employment toward skilled work within industries in *both* countries. This also offers an explanation of why trade could lead to increased relative demand for skilled workers within industries, rather than across industries, as predicted by the traditional HOS theory. Chusseau et al. (2008) and Pavcnik (2011) provide a summary of recent approaches of theoretical outsourcing models.

Various studies have tested the outsourcing hypothesis for single countries. Feenstra and Hanson (1996) found that outsourcing can account for a sizeable share of the increase in the relative demand for skilled workers in manufacturing sectors and for a notable amount of the increase in the relative wage of nonproduction workers in the United States during the 1980s.[35] Using updated data for the United States and measuring outsourcing by intermediate inputs in total materials purchase, Feenstra and Hanson (2003) found that outsourcing can account for half or more of the observed skill upgrading; the other half is contributed by technological change. For the United Kingdom, Hijzen (2007) also found international outsourcing contributing to the increase in wage inequality during the 1990s, although not to the same extent as technological change. Kang and Yun (2008) identified deindustrialization and outsourcing to China as two of the factors of rapidly increasing wage inequality in Korea since the mid-1990s, in addition to human

[34] The definition of outsourcing as "imports of intermediate inputs by domestic firms" is broader than the pure subcontracted part of the production process usually associated with outsourcing (see Chusseau et al., 2008).

[35] Feenstra and Hanson (1999) estimate that outsourcing could explain between 15% and 40% of the increase in wage inequality, depending on the specification.

capital factors and technological change. On the other hand, Slaughter (2000) suggested that outsourcing activities of US multinational enterprises tend to have small, imprecisely estimated effects on US relative labor demand. Similarly, using industrial data for a group of OECD countries, the OECD (2007) also concluded that outsourcing in general has only a rather moderate effect on shifting relative demand away from low-skill workers within the same industry. Lorentowicz et al. (2005), on the other extreme, discovered that outsourcing actually lowered the skill premium in Austria, a skill-abundant country, whereas it increased the wage gap in Poland, a relatively labor-abundant country.[36]

There are, however, few larger cross-county studies that explicitly test the outsourcing hypothesis. Taking outward FDI as a partial proxy for outsourcing, the OECD (2011) found this effect to be only modestly significant for explaining increased wage inequality in a sample of 23 OECD countries and distribution neutral in terms of overall earnings inequality (i.e., when employment effects are included).[37] This result is consistent with the fact that outsourcing activities to developing economies account for a small portion of total outward FDI stock in most OECD countries.[38] Analyzing 16 OECD countries over 1980–2000, Mahler (2004) also found that outward FDI is not significantly related to both household earnings and income inequality in either direction.

19.5.2.5 Technological Change

Next to trade and financial globalization, there are other equally plausible and competing explanations for income distributional changes. One that is often portrayed as an alternative to trade-related explanations is technological progress (e.g., Autor et al., 1998; Berman et al., 1998). Technological change, often described as advances in information and communication technology, is considered skill-biased insofar as it increases the total relative demand for skills for given prices of skilled and unskilled labor. Whether factor- or sector-biased (or indirectly biased via other factors of production), skill-biased technological change (SBTC) tends to increase the wage premium and/or increase unemployment among low-skilled workers and is therefore expected to increase inequality.[39] The wage premium will not increase only if the increase in the relative

[36] Some country-specific studies analyze the outsourcing effects on wage dispersion at the firm level. Analyzing data for the United States from 1981 to 2006, Ebenstein et al. (2009) suggested that the location of off-shoring activities matter, and off-shoring to high-wage countries can increase wages (via proliferation of nonroutine tasks), whereas off-shoring to low-wage countries have a negative wage effect.

[37] The same study also tested whether outward FDI has different effects in countries with distinct institutional settings (notably EPL), and found that outsourcing plays a modest role in wage inequality trends regardless of the institutional setting of the country considered.

[38] Intra-OECD investment, in fact, accounts for >75% of total outward FDI stocks in more than half of OECD countries (OECD, 2005).

[39] For the specific subset of central and eastern European transition countries, Vecernik (2010) suggests that differences in wages between skilled and unskilled labor were one of the major determinants of inequality increase after the economic transition in 1989.

demand for skilled labor is offset by a corresponding increase in the endowment with skilled labor.

In most studies, skill bias is identified by looking at changes in the share of skilled workers in sectoral wage bills or employment, and an increase in these shares within selected and defined research and development (R&D) industries or firms often is interpreted as evidence for SBTC.[40] Research that uses direct measures for technological progress such as computer usage or total factor productivity also reaches similar conclusions, although there is still debate over whether it is sector bias or skill bias that determines changes in the wage distribution.[41] The impact of technology seemed to be robust even when broader levels of aggregation were analyzed.

One reason why technological change often has been privileged over trade as the main explanation for increased inequality is the observation that employment shifts toward skilled work happening *within* rather than between sectors (although newer trade theories take this phenomenon into account in the frame of heterogeneity of firms models; see Section 19.5.2.1). Although this finding was confirmed for a sample of 12 OECD countries by the OECD (2011, p. 139), the analysis also highlights the growing wage inequality among workers with *similar* skills. Even after accounting for observable differences across workers, the dispersion of wages has risen, that is, there has been an increase in residual wage variation. The simple distinction between skilled and unskilled workers may not be detailed enough, and technological change, in particular ICT developments, can be accompanied by shifts away from routine and toward nonroutine labor (Autor et al., 2003; Michaels et al., 2010; Goos and Manning, 2007).

Many studies that have put technological change in the forefront of their explanation refer to one single country. Over the years, considerable evidence has been collected for the United Kingdom (e.g., Haskel and Slaughter, 1999; Hijzen, 2007) or for the United States (e.g., Blackburn and Bloom, 1987; Acemoglu, 1988; Card and DiNardo, 2002; Autor et al., 2003; Wheeler, 2005).

Larger cross-country studies including measures of technological progress (usually among the controls) became available more recently. Some studies identified this process as a key driver for inequality: the IMF (2007) finds that, overall (i.e., for the total sample of 51 countries), "technological progress has had a greater impact than globalization on (income) inequality within countries" (p. 31). Looking at the subsample of advanced countries, it turns out that globalization in terms of FDI contributed as much as, if not somewhat more than, technological change to increasing overall income inequality.

[40] Machin and Van Reenen (1998) and Autor et al. (1998) showed that such an indirect technology measure (i.e., the share of wage bills or employment) is highly correlated with direct measures of technological changes such as R&D intensity or computers.

[41] Krueger (1993), for instance, measures technology by computer usage, whereas Hijzen (2007) uses total factor productivity growth for skill-biased technical change. For a discussion of sector versus factor bias, see Haskel and Slaughter (2001, 2002).

A higher share of ICT investment also is identified as being strongly and significantly associated with higher inequality in 16 advanced countries by the ILO (2008).

The OECD (2011) also shows a strong and positive effect of technological change (captured by R&D business sector expenditures) on both wage dispersion among workers and overall earnings inequality among the whole working-age population. The second effect arises because technological change had no significant effect on employment rates, and the overall effect was therefore driven by the increased wage dispersion effect. Technological change is further shown to affect mostly the upper part of the distribution (OECD, 2011).

It is, however, in practice extremely difficult to disentangle technological change from other aspects of globalization that increase skill premia.[42] Advances in technology are, for instance, at the origin of the fragmentation of economic activities, outsourcing and off-shoring, or, as Freeman (2009) put it, "offshoring and digitalization go together."

19.5.2.6 Trade-Induced Technological Change or Technology-Induced Trade?

In most studies, technological change is treated as an exogenous variable (e.g., IMF, 2007; ILO, 2008; OECD, 2011). However, developments of technology and trade are not independent. Increased trade openness has contributed to the spread of technology, whereas technological progress has helped widen trade integration. Therefore, the three studies mentioned above recognize that technological change can also be seen as an additional channel through which economic globalization operates.[43,44]

Chusseau et al. (2008) reviewed four studies from the early 2000s, all of which found indications of trade-induced technological change in advanced countries. More recent studies confirm this picture. Bloom et al. (2011) showed that trade with low-wage countries (in particular China) had large effects on technical change in 20 European countries and the United States; it led to within-firm technology upgrading as well as between-firm reallocation of jobs towards more technology-intensive enterprises. Equally, Goldberg and Pavcnik (2007), Verhoogen (2008), and van Reenen (2011) emphasized in their studies that increased trade integration leads to faster technology upgrading.

Another approach to the interaction between globalization and technology has been called "defensive innovation" and goes back to Wood (1994). Firms that faced intensified import competition from developing countries have incentives to engage in more R&D efforts to develop new ways of production to remain competitive. While testing this

[42] As Wood (1998) argues for the period between the mid-19670s and mid-1990s, "there is plenty of evidence that skill-biased technical change has raised the relative demand for skilled workers, but much less evidence of an *autonomous* acceleration in its pace over the past two decades" (p. 1478).

[43] As Feenstra and Hanson (2003) put it, "Distinguishing whether the change in wages is due to international trade, or technological change, is fundamentally an empirical rather than a theoretical question" (p. 148).

[44] Institutions-induced technological change also has been proposed (see Chusseau and Dumont, 2012).

hypothesis is complex because it requires the availability of innovation data at the firm level, there are some studies confirming such an effect.[45]

The hypotheses of trade-induced skill-biased technological change SBTC and endogenous SBTC through capital deepening is also backed up by the OECD (2011), which suggested a positive correlation between SBTC, trade and capital flows, pointing to an interplay between globalization and technological change.

19.5.2.7 Education

Access to education and human capital accumulation are important factors that are expected to have an impact on income distribution. A higher average level of education is often expected *ceteris paribus* to reduce income inequality because it allows a greater share of the population to benefit from higher-skill activities (see, e.g., results from Sylwester, 2003 for OECD countries and an enlarged country sample for the period 1970–1990). However, while there is agreement on the existence of positive economic returns on education in terms of earnings levels, the theoretical predictions of the *inequality* effect of changes in education enrolment are not straightforward. Increases in education levels entail both a composition and a wage effect, which can move in different directions: the composition effect increases the share of higher education and initially tends to increase inequality before eventually decreasing it when higher education becomes the majority choice. The wage effect lowers the wage premium as the supply of more highly educated workers increases and thereby decreases inequality (for a discussion, see Bergh and Fink, 2008 or De Gregorio and Lee, 2002).

The important point to retain here is that the education–inequality relationship is neither monotonic nor linear, and the education effect can first be disequalizing and then equalizing, in analogy with the Kuznets process (see also Rehme, 2007). Further, there remains the issue of lagged reversed causality, with inequality levels at time t affecting education enrolment at time $t+1$.

Human capital can be seen as a complement to technology. Increases in human capital and in the supply of skills are necessary to decrease and eventually reverse the pressure to higher inequality that stems from technological change. The underlying logic is that technological change in the economy drives up the demand for higher-skilled workers, while the overall effect on inequality by and large depends on how elastic the higher education output is in relation to the increased demand. If the response is slow or inadequate, the skill premium of the more highly educated (the incumbent and the inflow as well)

[45] Thoenig and Verdier (2003) found support for defensive innovations by looking at the correlation between foreign competition and the share of skilled workers within the firm. Bloom et al. (2011) used technology data at the establishment/firm level for advanced countries and found that Chinese import competition has led to a considerable technological upgrading in European firms through both fast diffusion and innovation. They also showed that both Chinese imports and information technology intensity, in turn, are associated with an increase in the wage share of skilled workers.

increase, implying, by definition, an increase in inequality in a dimension (education) that plays a large role in explaining overall inequality (on this latter relationship see Ballarino et al., 2014). Such a view refers to the model of a "race between technology and education" going back to Tinbergen (1975).[46]

In many of the studies reviewed here, some education variable (e.g., share of adults with secondary or higher education, average school years) is introduced, most often as a control variable to capture human capital development. None of these studies suggest a positive association with inequality, that is, a disequalizing effect of education on earnings or income inequality but in their majority rather an equalizing one. This is particularly the case when the country sample is restricted to the OECD/EU area, and significant coefficients are reported, for instance, by the ILO (2008), OECD (2011), Afonso et al. (2010) and Cassette et al. (2012), as well as Cornia (2012) for Latin American countries. In terms of magnitude, according to the OECD (2011), the growth in average educational attainment over the 1980–2008 period offset to a great extent the disequalizing effect brought on by other factors, in particular SBTC. De Gregorio and Lee (2002), in one of the studies that specify educational factors—attainment and distribution of education—as the main explanatory variable in their models, suggest that these explain some but by no means all of the variation in income inequality across countries and over time. Nonetheless, their analysis confirms a negative relationship between income inequality and higher educational attainment (and a positive one with educational inequality) for a larger sample of around 60 countries.

On the other hand, the IMF (2007) suggests that there is an insignificant association between education and income inequality for both the OECD and an enlarged country sample. Carter (2007) and Bergh and Nilson (2010) even report a positive association, but their studies pool a subset of OECD with a larger number of mostly low-income countries. The point that a more highly educated labor force can contribute to greater income inequality in developing and emerging economies is also made by Carnoy (2011). This is related to increasing returns to university relative to secondary and lower education; decreasing public spending differences between higher and lower education; and increasing differentiation of spending among higher education institutions, with declining spending towards mass universities relative to elite universities.[47]

For the sample of OECD/EU countries, however, it is fair to say that most empirical evidence points to an equalizing effect of educational expansion. These results are also important for policy considerations drawn from cross-country studies of the multiple causes of inequality. If "up-skilling" of the population can indeed provide a most powerful

[46] A note of caution is warranted here. While appealing, such a model should not be applied mechanically because it does not take into account dynamics and ignores the interaction with the capital market (Atkinson, 2008).

[47] Carnoy (2011) underlines that some of these features also hold for the United States.

element for countering the trend towards increasing inequality, policy responses that focus on increased access to education will be more promising than those that concentrate on limiting economic globalization (and technological progress). They potentially have a double dividend by contributing to capturing benefits from increased economic integration and by keeping inequality levels lower or actually lowering them (see also Machin, 2009).

19.5.2.8 Going Beyond the Economic Notion of Globalization

Some authors have argued that the pure economic aspects of increased openness—trade, capital flows, foreign investment and so on—do not reflect the whole reality of globalization. Other more social, political and cultural aspects would also merit consideration (e.g., Dreher and Gaston, 2008; Zhou et al., 2011; Atif et al., 2012; Heshmati, 2004). These authors typically construct synthetic measures of globalization along the lines of the Kearney globalization indexes[48] and test their significance and that of their subcomponents for explaining earnings and income inequality.

Interestingly, some of these studies—in particular Heshmati (2004) and Zhou et al. (2011)—find overall globalization to have a *negative* relationship with income inequality.[49] In these cases, investigation of the subcomponents of globalization reveals that the economic aspects (such as trade) tend to have a significant positive relationship, which is, however, more than outweighed by factors such as increased personal contacts/travel and information/Internet use.

While the above two studies of the impact of "overall" globalization are based on a broad country sample of advanced and developing countries (60 and 62, respectively), the Dreher and Gaston (2008) study allows the OECD area to be separated out in their analysis of 100 countries. For the OECD sample, they found overall globalization to have a significant *positive* relationship with inequality, whereby this association is much larger for earnings than for income inequality.[50] Different than the studies mentioned above, the three subdimensions of globalization (economic, political, social) seem to have no systematic relationship with inequality except that none of them have a negative sign in any of the specifications. Bergh and Nilson (2010) are another example of an analysis of the effect of an overall indicator of globalization and its element on net income inequality trends over the past 35 years in around 80 countries. Their results reveal a

[48] The Kearney Globalization Index (KGI) (see Kearney, A.T., Inc. and the Carnegie Endowment for International Peace, 2004, 2007) is composed of four major component variables: economic integration, personal contact, technological connections and political engagement. Each of these four component variables is a weighted average of several determinant variables. In a similar vein, Dreher (2006) proposed a composite measure for 123 countries, the KOF index of globalization, which is based on 23 variables that relate to three globalization dimensions: economic integration, political engagement and social globalization (see http://globalization.kof.ethz.ch/).

[49] But see a critical review of their methods and results in Atif et al. (2012) and Tsai et al. (2012).

[50] They estimate that a one-point increase in the overall globalization index increases industrial wage inequality by 26% and household income inequality by 3%.

positive and strong association[51] that is largely driven by the social dimension of globalization. Although the sign and size of the economic and the political dimensions of globalization are similar, their coefficient is not significant.

19.5.3 Changes in Institutions and Regulations

Until 30 years ago, the quest for identifying driving factors of income inequality focused on testing the Kuznet hypothesis (see Section 19.5.1). However, since the 1990s a range of other factors has increasingly been considered. In the context of OECD countries, globalization and technological change became prime candidates for research (many other variables show little variability in the OECD). It is, however, important to also consider the role of institutions, in particular labor market institutions, and changes in regulations (Checchi and Garcia-Penalosa, 2005; Piketty and Saez, 2006; Lemieux, 2008). The increase in wage inequality since the 1980s in several countries coincided with changes in labor market institutions, such as a decline in the importance of unions in setting wages. That labor market institutions and policies have lost redistributive potential in recent times also has been put forward; in particular, trade union density, collective bargaining coverage and centralized collective bargaining were estimated to have become less effective in reducing inequality (Baccaro, 2008). Chapter 18 provides a detailed discussion of the theory and literature that relates labor market institutions to the dispersion of wage earnings and proposes an empirical approach for analysis.

While it is widely recognized that institutions are an important factor for identifying the multiple causes of inequality (e.g., Acemoglu, 2003; Smeeding, 2002), the weight attached to this factor in econometric studies has long been limited. Some papers have argued that, given the relative stability of institutional patterns across countries, including country fixed effects in the analysis would capture a larger part of this factor, at least its time invariant components (e.g., Figini and Görg, 2006). This does not, however, fully reflect development over the past decades, during which some institutions such as union density and coverage or EPL considerably weakened in many countries.

In the earlier studies, the degree of unionization was the main factor used to measure labor market institutions (e.g., Freeman, 1993); union density (share of employees who are members of a trade union) or union coverage (share of employees covered by wage bargaining agreements) are probably more precise indicators. Union density and coverage often are expected to have an equalizing effect on the earnings distribution, not only because unions strive for wage standardization and seek to increase the earnings of their members[52] but also through indirect effects, such as promotion of social expenditures that

[51] Their results suggest that the maximum effect of overall globalization would be a 14% increase of the Gini coefficient of income inequality.

[52] The existence of wage premia for union members tends to be equalising if low-wage earners were better organized than high-wage earners, but the opposite may hold if high-paid earners were better organized (Freeman, 1993). Blau and Kahn (2009) argue that the net effect of unions on wage inequality partly depends on which groups have higher labor demand and supply elasticities.

benefit low-income groups as a whole (Mahler, 2004), creation of an institutional environment in which workers care more about wage dispersion because of some shared norm of fairness (Golden and Wallerstein, 2011) or employers following certain pay norms where workers are paid a fraction of their productivity plus a uniform amount (for a discussion of this reputational approach see Atkinson, 2002).

Another factor increasingly analyzed is the impact of wage-setting centralization and coordination. Again, this factor may have both direct and indirect effects on the distribution of earnings: centralized bargaining improves the bargaining position of workers; it may help broaden norms of distributive justice; and it is expected to be economically more efficient, resulting in more resources to be distributed (Mahler, 2004; see also the discussion in sub-section below).

A third factor that is expected to have an important effect on wage dispersion is EPL. EPL is likely to affect employers' costs to hire/dismiss workers. Such policies would compress the wage differential if they are relatively more important for unskilled workers. There may, however, be considerable differences for the effects of changes in EPL for regular versus temporary workers.

Further, there are a number of regulative factors that affect the distribution of earnings, such as minimum wages, unemployment benefits and tax wedges. The working hypothesis here is that minimum wages compress the wage differential, and a decrease in minimum wages contributes to an increase in wage inequality. Higher unemployment benefit replacement rates would increase the reservation wage, with a possible equalizing effect on wage inequality. The distributive effect of tax wedges is a priori ambiguous. Finally, not only labor market institutions and regulations affect the earnings distribution; the observed trend of a large decline in product market regulation (PMR), which precedes the larger trends weakening labor market institutions, also is expected to have a major role (OECD, 2011).

Many of the above aspects of labor market institutions and regulations are, in general, expected to have a more or less equalizing effect on the distribution of wages. This is, however, not necessarily the case when it comes to household earnings or income inequality; the latter also is influenced by trends in employment and unemployment at the household level. Rising employment, for instance, may attenuate growing wage inequality, and the net effect of institutions on household income inequality also depends on their effect on employment. A vast body of empirical evidence points to a significant effect of both institutions and regulations on employment levels (for an overview, see OECD, 2006).[53] Theoretically, the overall impact of institutions and regulations remains ambiguous (Checchi and Garcia-Penalosa, 2008).

[53] For evidence on unemployment benefits, see, for instance, Nickell (1998) and Nunziata (2002). For evidence on labor market bargaining models, see Layard et al. (1991) or Pissarides (1990). For evidence on product market regulation, see Blanchard and Giavazzi (2003), Spector (2004), Messina (2003), or Fiori et al. (2007).

The majority of studies reviewed (with the major exception of ILO, 2008) point to a negative association between various aspects of institutional and regulatory change and earnings as well as income inequality. Weakening of institutions has often been identified as a key driver of increasing inequalities.

19.5.3.1 Wage Dispersion Effects

Earlier studies of single OECD countries found that the decline in unionization increased wage inequality (Card, 1996; Machin, 1997). Looking at trends in a cross-country setting up to 1995, Rueda and Pontusson (2000) suggested higher union density is associated with a more compressed wage dispersion independent of the policy "regime" of a country (social, liberal, mixed). For the same set of OECD countries, Golden and Wallerstein (2011) provide newer estimates but make a distinction between the 1980s and the 1990s: in the former decade, decreasing union density and centralization were identified as key factors of increasing wage dispersion, whereas these factors were no longer significant in the 1990s and were replaced by trade and social expenditures as explanatory factors. Cassette et al. (2012) found union density and union concentration to be significantly negatively associated with earnings inequality for a set of 10 countries for a period of 25 years (up to 2005). Such a finding is also reported by Burniaux et al. (2006), although it is limited to particular inequality indexes. On the other hand, Mahler (2004) founds no effect of union density but a significant and negative effect of wage coordination on earnings inequality for a set of 13 OECD countries over the two decades 1980–2000.

Koeninger et al. (2007) found changes in a set of labor market institutions explained as much as trade and technology: EPL, levels and duration of benefit replacement rates, union density and the minimum wage were shown to negatively affect the wage differential. Checchi and Garcia-Penalosa (2005) identified three types of labor market institutions as essential determinants of wage differentials: union density, the unemployment benefit and the minimum wage. Declining minimum wages also have been found to increase wage dispersion, mainly at the lower end of the distribution (Dickens et al., 1999; DiNardo et al., 1996; Lee, 1999).

The OECD (2011) considers a range of labor market institutions and regulations as possible explanatory factors for increased earnings inequality in 23 OECD countries up to 2008. The weakening in these institutions and regulations since the 1980s was shown to widen the wage dispersion among workers: (i) the effect of EPL is entirely driven by weakening EPL for temporary workers, whereas EPL for regular workers had no significant effect. Furthermore, EPL had more of an impact on the lower than the upper half of the earnings distribution; (ii) lower unemployment benefit replacement rates for low-wage workers (but not for average-wage workers); (iii) decreases in union coverage, which predominantly affected the upper half of the earnings distribution; and (iv) and lower taxation of earnings (tax wedge).

Effects of changes in product market regulation are generally not included in analyses of inequality but rather are considered in studies of employment effects (e.g., Nicoletti and Scarpetta 2005; Bassanini and Duval, 2006; Fiori et al., 2007). However, it can be expected that these regulations had a larger role in wage dispersion. The OECD (2011) showed that declining PMR contributed significantly to a wider wage dispersion, in particular at the lower half of it. This is consistent with the view that PMR tends to reduce market rents available for unions to capture through collective bargaining (Nicoletti et al., 2001); this leads to a decline in union power (or more decentralized bargaining), which in turn results in greater wage dispersion.

Combining the results of the effect of institutions on wage dispersion with additional ones on employment, the OECD (2011) estimated the overall effects on earnings distribution among the entire working-age population. It turns out that wage dispersion and employment effects often were off-setting and led to undetermined estimates of the effects of institutions and regulations on overall earnings inequality, with one exception: weaker employment protection among temporary workers, which is estimated to have an overall disequalizing effect.

19.5.3.2 Income Inequality Effects

Some studies provide estimates of the direct effect of institutions on (gross or net) income inequality, in particular Checchi and Garcia-Penalosa (2005, 2008) and the ILO (2008). All three studies cover a set of 16 OECD countries for a period up to the early 2000s. Checchi and Garcia-Penalosa (2005) identify union density, the tax wedge and unemployment benefits as major determinants of higher income inequality, whereas the effect of minimum wages is only marginally significant. The overall effect of stronger institutions is estimated to reduce income inequality, partly through wage compression and partly through a reduction in the rewards for capital. For a smaller sample of seven OECD countries, Weeks (2005) estimated decreasing union density as a strong predictor of increased gross income inequality.

Based on a different set of data that allows several income concepts to be investigated, Checchi and Garcia-Penalosa (2008) suggested only a weak role for institutions in determining factor income inequality. A stronger effect occurs when considering disposable income inequality, particularly for unemployment benefits and EPL (negative) as well as tax wedge (positive), whereas union density, wage coordination and minimum wage remain insignificant. The fact that the tax wedge is estimated to increase income inequality (including factor income inequality) runs counter to some of the evidence summarized earlier. Checchi and Garcia-Penalosa (2008) put forward that high-wage workers may be better able to pass tax increases onto their employers than low-wage workers and that a high tax wedge can increase unemployment.

Results reported by the ILO (2008), based on Baccaro (2008), show that trade unionism and collective bargaining are not significantly associated with within-country inequality, except in the central and eastern European countries.[54] Rather, economic factors such as technology-induced shifts in the demand for skilled labor and increases in FDI shares seem better predictors if increasing inequality. This nonsignificance of institutional factors also holds for the enlarged sample of 51 countries going beyond the subsample of the 16 OECD countries. Evidence for 14 OECD countries, presented by Mahler (2004), is quite the opposite: union density and wage coordination were found to have the strongest negative relationship with disposable income inequality, whereas indicators of economic globalization (imports, outbound investment, financial openness) were found to be insignificant.

19.5.4 Political Processes

A great deal of the political science and of the policy literature is concerned with the effects of inequalities and how they can be mitigated in various societies. For this chapter, however, it is the other direction that is interesting: mechanisms of how various political arrangements (voting, electoral institutions and representation in political parties, interest reconciliation and employer–employee relationships) affect inequality. The core question is, therefore, How and to what extent can political factors account for the variability of inequalities across countries and over time? How much of the cross-country and over-time variance of inequality can be explained by political determinants (agency,[55] institutions or policies)?

The explanation of inequalities by political institutions has to start from the actual level and structure of inequality itself (initial or t_1 distribution). Then the degree of change achieved by institutions and policies—how they modify the social setting and transform it into a new system of inequality (end result or t_2 distribution)—is subject to study here. The assumption is that the objective position in the income distribution defines preferences over redistribution, which is aggregated in the political process, the end of which, in turn, is a change in income distribution. This is, no question, a loop in the line of reasoning, indicating a circularity in the arguments. This is a difficult issue for empirical research and, although recognized by many, few have offered convincing solutions to it.

We classify the channels of this transformation into three groups: (i) democratic representation and partisan politics, (ii) interest groups and lobby organizations and (iii) redistributive policies of the state (governments). From a different angle, we are concerned with the demand for and the supply of policies, mediated by the political process itself. Below we turn to these in detail.

[54] Bradley et al. (2003) also report the "absence of any significant effect of wage coordination on pre-tax and transfer inequality" (p. 216) for the 61 countries they investigated.

[55] There is no question that agency (political leadership) may exert influence on the shape of inequality, especially for shorter periods and especially in countries where the political system allows for a larger role of personalities. This happens in fully democratic states, less democratic states and nondemocratic environments, in "normal" democracies and in populist regimes. Nevertheless, dealing with the role of political personalities would stretch beyond the scope of this chapter.

19.5.4.1 Preference Formation and Partisanship

19.5.4.1.1 General Frame of Understanding

The most commonly used general frame for understanding the politics of redistribution in democratic societies is offered by Meltzer and Richard (1981), originating from a Downsian definition of political competition and democracy (Downs, 1957; see also Romer, 1975). In this setting politics is about redistribution only, and the extent of redistribution is defined by electoral politics only. The aim of parties is to win elections. It is assumed that in majority voting systems (where the winner takes all) the party that is able to attract the vote of the median voter—the median being defined in terms of the dimension in which the political agenda stretches the political spectrum (incomes, political opinions, etc.)—wins. For voting on taxes and redistribution, the spectrum is, by definition, defined by the level of incomes/wealth. Voters, who by their material wealth/incomes occupy the full continuum of the income distribution, vote over the general tax rate, which provides resources (public funds) for redistribution. If the pivotal voter is the same as the person with a median income (which is not necessarily the case), on the assumption of self-interest he or she would prefer more redistribution (higher taxes) than a person with an income above the median. An increase in inequality can be gauged by the increased distance between the median and the average income. The demand for redistribution in period t_2, therefore, is assumed to be linked to the extent of inequalities in period t_1. Under the Meltzer and Richard (hereafter MR) paradigm, greater inequality leads to higher social spending and results in larger redistribution. This would imply a higher level of redistribution in countries with greater inequalities to start with. To put it differently, multiparty democracy, as described above, would produce an equalizing self-correction mechanism, leading to larger redistribution in those countries where inequalities are larger. The prediction, therefore, is that the variance of inequalities are, at least to some extent, dependent upon the essential features of democracy.

There have been many tests of this proposition, contrasting levels of inequality with levels of redistribution, with varying results. As an empirical test, for example, Milanovic (2000) found that there is a consistent association between gross household income inequality and more tax/transfer redistribution in a set of 24 democracies in the period of the mid-1970s to the mid-1990s. Also, Mahler (2008) found support for the MR propositions after refining definitions of original inequality and redistribution.[56] Mahler (2010) found a positive relationship between pregovernment inequality and government redistribution on the basis of observations of 13 OECD countries. Mohl and Pamp (2009) stated that there is a nonlinear relationship between the two. They concluded that at very high levels the positive relationship between inequality and redistribution is reversed. The argument for the reversal stresses the role of Director's law, that is, that redistribution

[56] When, however, it is not the status (democratic preference aggregation via representative democracy) but the process itself (say, transition from nondemocracy into democracy) that is observed, Nel (2005) did not find support for the median voter hypotheses (despite careful definitions of the variables used).

may go from the ends to the broadly defined middle class (ranging from the 20th to the 80th percentile).[57]

Contrary to the above findings, and partly because of lack of appropriate data or improper specifications, many of the tests of the link between initial inequality and redistribution could not reach conclusive results. (For reviews of various aspects of the MR model and its propositions, see Alesina and Giuliano, 2009; Borck, 2007; Guillaud, 2013; Keely and Tan, 2008; Kenworthy and McCall, 2007; Lübker, 2007; Lupu and Pontusson, 2011; McCarty and Pontusson, 2009; Mohl and Pamp, 2009; Olivera, 2014; Osberg et al., 2004; Senik, 2009.)

A potential reason for the inconclusiveness of the literature may be that, as Robinson (2009) put it, "The model does not predict a simple positive relationship between inequality and redistribution across countries since there are many differences between countries which may be correlated with either the demand or supply of redistribution at a particular level of inequality" (p. 28). Also, it can be expected that in high-inequality countries with badly performing institutions, any income that is taxed away is likely to be wasted by corruption or diverted by elites, and this will reduce the demand for redistribution. Also, in general, MR would mean that extension of the franchise will increase redistribution, that is, democratization of the political regimes brings about lower levels of inequalities. However, while the equalizing effects of democratization seem to be shown in many cases, they might not be automatic (see Galbraith, 2012; Nel, 2005; Robinson, 2009).[58]

In what follows we go through some relevant assumptions and predictions and use the MR proposition to structure the line of reasoning here, acknowledging the fact that some alternative suggested theoretical papers (most notably Iversen and Soskice, 2006 and to some extent Moene and Wallerstein, 2001) suggest different frames and sometimes diametrically different conclusions. We start from the micro (assumptions on the motivational base of voters) and move to the macro level (such as features of electoral systems).

[57] When referring to a "pregovernment" situation, one needs to keep in mind that the data relate to incomes before taxes and transfers in the presence of government. The "before redistribution" inequality is affected by the existence of the government, and it is quite possible that this is greater than the inequality that would be found if the government were not present.

[58] A more recent attempt to trace inequality paths among 30 developed societies points out that countries experiencing democratization in central and eastern Europe followed very different paths in terms of inequalities. While all belonged to the lower end of the inequality spectrum in the 1980s, they ended up at very different parts of the European "league table" in the late 2000s: Slovenia and the Czech and Slovak Republics at the bottom and the Baltic states on top, while the rest lie in between. The experiences of Spain, Portugal and Greece, where the ending of the dictatorships went hand in hand with inequality decreases, therefore, have to be balanced against the experiences of the central and eastern European countries in further comparative research (Tóth, 2014). However, a major difference between the Mediterranean and central and eastern European transitions was clearly that, in the latter group of countries, transition also implied marketization and liberalization, in contrast to countries in southern Europe where the role of the state changed albeit less in scope.

A simple presentation of the potential links between inequality, redistribution and intermediate processes is shown in Figure 19.2 (following Tóth et al., 2014). As indicated in Figure 19.2, there are potential mediating mechanisms on both the micro and the macro levels. On the one hand, personal attributes and perceptions might have an effect on individual redistributive preferences and, on the other, the institutional mechanisms that translate preferences to policy actions. Determinants of political participation shape the ratio and the composition of voters, and the activity of the civil society matters a lot in policy decisions. Finally, it is clear that the ways in which (and to what extent) attitudes of voters will, via the machinery of politics, shape policies depend to a large extent on various institutions (political and executive alike).

19.5.4.1.2 Motivations, Expectations and Values of Voters

To understand the mechanisms of the micro determinants of votes over redistribution is crucial and has to be linked more closely to the political science literature. However, a large number of empirical studies are already available and provide more understanding of the characteristics and *motivations* (from the redistribution perspective) of citizens belonging to various parts of the income distribution. Various studies show that although it exists, the correlation linking material position and attitudes regarding the welfare interventions of the state is far from perfect. Some attempts to identify reasons for the

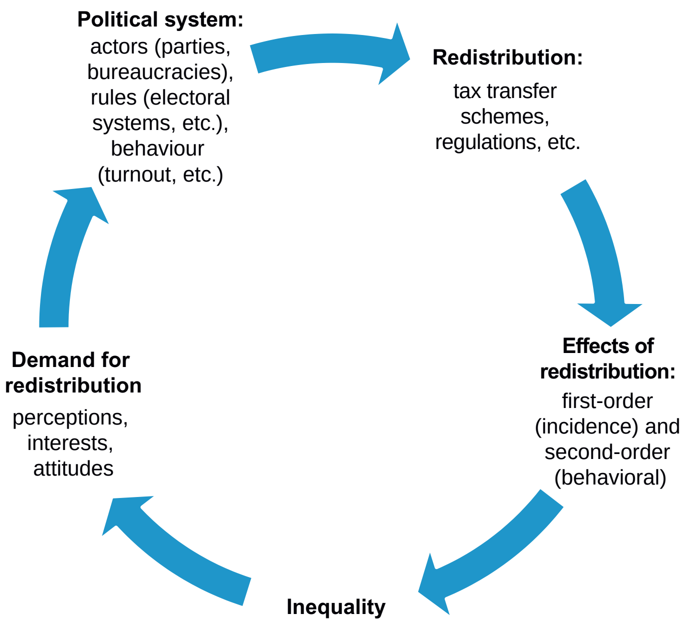


Figure 19.2 Theoretical links of the political processes involved in the determination of income distribution. *Source: Tóth et al. (2014)*

"deviations" (i.e., the observation that some of the relatively richer voters will be pro-redistribution while others with below-median incomes may not be supportive) stress that it is not only the current economic position but also the expectations concerning economic prospects that matter (see Bénabou and Ok, 2001 and Ravallion and Loskhin, 2000 for prospect for upward mobility; see Alesina and Fuchs-Schündeln, 2005, 2005, Piketty, 1995 or Guillaud, 2013 for social mobility experiences and expectations based on these[59]).

Others stress the role of socialization into general *value systems* either in the frame of the overall sociopolitical environment, such as a socialist past, or simply ideological systems or family traditions (Kelley and Zagorski, 2004; Corneo and Grüner, 2002; Fong, 2001, 2006; Alesina and Fuchs-Schündeln, 2005; Gijsberts, 2002; Suhrcke, 2001). These are, in many cases, not temporary but long-lasting cultural differences, sometimes transmitted over generations (Alesina and Fuchs-Schündeln, 2005; Luttmer and Singhal, 2008). Also, the beliefs about the fairness of the economic system and about the rules of the game of "getting ahead" in society seem to be important determinants of the acceptance the actual level of redistribution or a demand for more of it (Fong, 2001, 2006; Alesina and La Ferrara, 2005; Alesina and Glaeser, 2006; Osberg and Smeeding, 2006; for a recent review of the literature on inequality and justice perceptions see Janmaat, 2013).

Finally, it is not simply general views and attitudes but also personality traits that can matter. A hypothesis of how these attitudes come about is presented by Tepe and Vanhuysse (2014). They found that personality traits in some cases strongly determine welfare attitudes, even after controlling for class, sociodemographic variables and even socialization.[60] Moreover, they show that some traits such as conscientiousness, openness and extraversion are conditioned by communist regime socialization (when comparing the Eastern and Western Länder of Germany, similar to Alesina and Fuchs-Schündeln, 2005).

19.5.4.1.3 Reference Groups and Heterogeneity of Voters

Inequality is often measured by various indices reflecting the whole income distribution (most commonly by the Gini coefficient but also by various other variance-based measures). Putting these into the right-hand side of regressions is, however, problematic in political economy models. It cannot be reasonably assumed that voters have the same image of inequality that is provided by any of these rather complicated measures. It is a much more plausible assumption that voters think of social distances, define proximity to other voters, etc. The idea of social affinity (an acknowledgement of those groups who

[59] As for the measurement of and trends in actual income (and social) mobility, Chapter 10 of this book provides an exhaustive overview.

[60] As an example, the research of Tepe and Vanhuysse (2014) suggests a positive relationship between agreeableness and support for the state's role when unemployed, a negative relationship between openness and support for governmental responsibility for the family and a positive association between conscientiousness and governmental responsibility for the elderly.

are the closest to the assumed decision makers) was raised by Kristov et al. (1992). For political economy models of redistribution the idea has been applied by Osberg et al. (2004), Lupu and Pontusson (2011), Finseraas (2008) and Tóth and Keller (2013). Empirical tests show that that the actual level of inequality (and, more importantly, the structure of inequality as measured by the distance between the middle classes and the poor) also drives attitudes towards redistribution. There seem to be convincing examples that the relative position of the middle—which might cover also the pivotal voter in elections—influences public spending priorities (and coalition formation). As Lupu and Pontusson (2011) showed, a greater dispersion in the lower half of the earnings distribution (as measured by the P50/P10 ratios) is consistently associated with less redistribution in a sample of 15 advanced democracies. A more prominent skew of the redistribution (meaning middle classes being positioned closer to the poor) would result in more redistribution in their sample. Osberg et al. (2004) also showed that the structure of redistribution matters, but in a different way: they found that inequality between the top and the middle of the distribution (measured by the 90/50 ratio) has a large and negative effect on social spending, implying that the top may have more room for opting out of public services in the case of larger inequalities.

19.5.4.2 The Issues at Stake: Different Forms of Redistribution

The assumption of the basic MR model is that there is only one type of redistribution (vertically transferring money from the rich to the poor). The original model is even more simplistic: it specifies a uniform tax rate levied on the above-average-income voters on the one hand and a lump sum amount handed over to the lower segments of the distribution. Actual redistribution programs are, however, more sophisticated. As Moene and Wallerstein (2001, 2003) pointed out, distinction between insurance-type programs (in which participants seek provisions against income losses at bad times) and redistribution programs involving taxes on the rich to benefit the poor has to be made. They suggest (and offer empirical evidence to support the suggestion) that while the demand for vertical redistribution is negatively correlated with income, the demand for insurance is positively correlated (and in some situations these two effects might even cancel out each other). This might indeed have a sizeable effect on the actual distributive outcomes.

In his review of the literature, Borck (2007) summarized various types of redistribution and classified the literature according to this differentiation. The first and most obvious direction is redistribution from the rich to the poor; models underlying social preferences, upward mobility and voter mobilization (see above) point to the direction of causation from increased inequality to increased vertical redistribution. There are, however, other types of redistributive mechanisms, such as spending programs, that entail transfers from the poor to the rich. This might be the case when there is public provision of private goods, education or insurance. In these cases the state/public budgets may effectively be subsidized by the poorer income groups. Finally, the public provision of private goods or the operations of public pension schemes might represent a case for the so-called Director's law: when the tails of the distribution are expropriated by the middle (for other reviews, see Mohl and Pamp, 2009; Mahler, 2010; Alesina and Giuliano, 2009).

Another issue regarding the definition of redistribution relates to the income concepts used for measurement. Obviously, simply associating Gini coefficients after taxes and benefits with the size of the public social budgets is erroneous because it conflates the right- and the left-hand sides of the equation. Based on LIS data, Kenworthy and Pontusson (2005) refined the definition of redistribution. They proxy redistribution by a difference between the Gini of disposable household incomes (after taxes and benefits) and the Gini for market incomes (before taxes and benefits). This helps them show (on both cross section and on country time series data) that an increase in market income inequality correlates with an increase in redistribution (see similar results from Immervoll and Richardson, 2011).[61] This finding about the over-time, within-country variation of redistribution as a response to inequality is in broad agreement with what is suggested by the MR proposition. What makes a difference between countries, however, is the elasticity with which the welfare states react during the period they observe (varying spells in the 1980s and 1990s) an inequality increase.[62]

An additional empirical characteristic of electoral politics is that sometimes parties do not simply play the cards of (vertical or insurance-type) redistribution in elections. They often try to make political space multidimensional, sometimes introducing issues that create divisions orthogonal to the vertical income differentiation. Campaigns often are about complex packages, and "issue bundling" might easily place the median voter at a part of the income distribution different from the median income (Roemer, 1998). This might, in concrete circumstances, be a strategy to target parties on the Right of the political spectrum (because they are interested in diverting the electorate away from issues that motivate the lower-income groups), but issue bundling may sometimes also be in the interest of Left parties.[63]

19.5.4.3 Political Inequality: Unequal Participation in Elections

The prediction of higher redistribution in the case of higher inequality also assumes full (or at least uniform across income groups) participation in elections. This, however,

[61] A special note is needed here. Increased redistributive effects of given welfare state measures may be detected induring periods of increasing market income inequality, even in the absence of any changes in redistribution instruments such as taxes and transfers. We turn back to this in Section 19.5.5.1. Also, see Immervoll and Richardson (2011).

[62] Also the choice of the country universe in this case can clearly make a difference in results. OECD comparisons (see, e.g., OECD, 2011) tend to show a great deal of sensitivity of the results to the inclusion of lower-income OECD countries (such as Mexico or Chile).

[63] A further analysis of issue bundling would reveal how politics and policies that are not directly aiming inequalities could have important effects on actual developments in income distribution. This way of taking account of secondary effects, by-products and unintended consequences of party politics would, however, go beyond the scope of this chapter.

generally does not hold empirically.[64] Therefore, differential voter participation might alter aggregate redistributive preferences. If the middle classes participate more than the poor, then parties may seek to represent the interests of relatively higher-income voters. In another dimension, greater participation of older voters can induce more party promises for pension expenditures compared with family-related expenditures. Therefore, empirics of the actual redistribution might differ from predictions based on uniform participation. (See more on participation in Kenworthy and Pontusson, 2005; Larcinese, 2007; Pontusson and Rueda, 2010.)

An important note by Kenworthy and Pontusson (2005) and, especially, by Pontusson and Rueda (2010) is that the mobilization of voters is a crucial issue in how inequality translates into politics of redistribution. Political inequality (at least in terms of participation in elections) may play a major role in policy formation. Because the low-income voters who might be motivated in larger redistribution may not be sufficiently activated during elections, redistribution might be lower than predicted by "objective" inequality. Pontusson and Rueda (2010) also point out that there is a need to differentiate between core constituencies of the Left (and Right) parties, in addition to the positions of the median voters who, in proportional representative (PR) systems at least, can be considered swing voters. Their major finding is that the extent to which Left parties take up the issue of redistribution also depends on the general mobilization of low-income citizens. To put it differently: if the "demand" for redistribution is represented by a larger appearance of the low-income segments in the polls, the Left will react to it by offering more redistributive policies. This, of course, cannot fully be treated as exogenous; therefore, party politics for differential mobilization of their core constituencies (especially on the Left) might have an important effect on redistribution. This issue is discussed further in the next section (Section 19.5.4.4) on political institutions.

Mahler (2008) introduces two factors into the analysis: the level of electoral turnout and the degree to which turnout is skewed by income. When these factors are taken into account, the predictive power of the MR model is significantly improved. He found the link to be especially strong for the lower and the middle parts of the income distribution and when social transfer policies are at stake as opposed to tax policies. In a later and more refined formulation, Mahler and Jesuit (2013) showed that political participation (most notably union density) is positively related to redistribution, especially when the share gains of the lower middle classes are considered.

[64] Full participation should not even be assumed theoretically. Following Downs (1957) and Olson (1965), it is shown and accepted in mainstream political economy thinking that voters are perfectly rational not to participate in elections, while it is also rational for rich voters/small interest groups to lobby and fund parties (see Olson, 1965). For an overview of what political economy reasons can be found behind insufficient performance MR-type and other "economistic" approaches to politics that work via "the market for votes" analogies, see Vanhuysse (2002).

19.5.4.4 Political Regimes and Partisanship

For a broader understanding of the effect of political dynamics on income distributions, it is worth starting with a consideration of the effect of general political regimes—most notably democracy—on inequality. As stated by Galbraith (2012) in a review of many propositions, it is difficult to establish clear conclusions. Classifying political regimes into democracies and nondemocracies does not help much. Some nondemocratic (communist or Islamic) regimes can have more egalitarian distributions than others. Of course, long-serving, established social democratic regimes of the twentieth century are associated with lower-level inequality, but causality may run in either direction. Finally, there are numerous examples when the transition to a more democratic regime is paralleled by an increase rather than a decrease of inequality (consider the case of central and eastern European countries experiencing post-communist transitions) (Galbraith, 2012; Tóth and Medgyesi, 2011; Tóth, 2014).

Second (and more generally), because various "welfare regimes" (the term coined by Esping-Andersen, 1990 in classifying the overall characteristics of the European welfare systems into three types of welfare regimes à 'la Esping-Andersen) are so embedded in general socioeconomic and sociopolitical settings, partisanship (normally meaning parties staying in an executive position for one or two election terms) cannot really achieve fundamental changes in the operation of an overarching institutional setting. Both of these considerations lead us to an analysis of not only the general frames of the political regimes, such as representative democracy, but also to elements of these (such as partisanship, ideologies, corporatist institutional settings).

A large tradition of the political science literature associates redistribution to the relative strength of the parties representing the working class in elections. Social democratic parties have long governed some democracies with large public spending, although their socioeconomic foundations have declined with the large sectoral shifts in economies following the two consecutive crises in the 1970s. However, the power resources theory (PRT) is an influential paradigm in explaining redistribution, arguing that the extension of the welfare state largely depends on the ability of the parties representing labor to mobilise lower-income voters (Korpi, 2006; Korpi and Palme, 2003).

Bradley et al. (2003), using a panel of 19 OECD countries, attempt to explain what determines "initial" income distribution and what are the results of redistribution and provide support for the central hypotheses of the PRT. They stress that high unemployment, low union density and a large proportion of households led by women are associated with high inequality before taxes and transfer. For the reduction of inequality (i.e., the effectiveness of the welfare state redistribution) they identify the existence of Leftist government (either directly or indirectly via other variables related to partisan politics) as statistically significant (and strong). As they conclude, "leftist government very strongly drives the redistributive process directly by shaping the redistributive contours of taxes and transfers and indirectly by increasing the proportion of GDP devoted to taxes and transfers."

Iversen and Soskice (2006) allow for heterogeneity of parties (assuming separate, exclusive representatives of high-, the middle- and the low-income voters). They also allow coalitions between the representative parties, and they differentiate between PR and simple majoritarian electoral systems. Their proposition is that majoritarian systems tend to redistribute less because they tend to favor centre-Right governments (as a result of the interplay of the coalition game under constraints of the potential taxability of the three major income groups). Note, however, that Iversen and Soskice (2006) do not build on assumptions about the relationships of the mean and the median incomes (i.e., about the level of inequalities in the society), nor about the position of the median voter in the income distribution (i.e., about the effect of political mobilization on political coverage of the full income spectrum). Their assumed parties are, however, class parties representing the various income groups. The core element of the argument is the nontaxability of high-income groups and the uncertainty about the potential to enforce pre-election party commitments after a coalition is formed.

Some empirical accounts of the political dynamics and its effects on inequality challenge the usefulness of the traditional notions of Left–Right differentiation, and they also add to a more balanced understanding of the meaning of various "regimes." As Rueda (2008), for example, stresses in his study of 16 OECD countries, in regimes where the underlying socioeconomic structure is characterized by corporatism (a broad, concentrated, institutionalized and informal system of bargaining and interest reconciliation between social partners, state bureaucracies and political parties), a small part of the discretion over, for instance, wage policies remains in the hands of partisan politics—hence the nonsignificance of the partisanship variables in explaining income distribution. In addition (as also put forward by Rueda, 2008), Left parties may (contrary to their general image) not always represent the full "labor side" of the economy. Rather, they may be more concerned with "insiders" (the employed, in this case) of the labor market rather than the "outsiders," who may wish to enter the labor market but are not (yet) there. With outsiders' interests being overlooked, inequality of overall incomes may increase even in periods of Left governments.

Rueda and Pontusson (2000) analyze four relevant political-institutional variables to explain (wage) distribution in a set of advances countries[65]: in addition to the partisan composition of government, they measure unionization rates, centralization of wage bargaining and the size of the public sector. They observe the effect of these variables in two different broad institutional contexts: social market economies (SMEs) and in liberal market economies (LMEs), as defined by Hall and Soskice (2001). The former setting is characterized by comprehensive, publicly funded welfare systems, heavily regulated labor markets and institutionalized wage bargaining systems. They find that these two distinctive general settings do have an effect on wage formation and distribution. Except for

[65] To account for broader socioeconomic variables, they control for participation of women in the labor force and unemployment rates.

unionization, for which the above broader institutional settings are not significant (higher unionization has an equalizing effect in both regimes), the effect of the other observed institutional variables differs in the various variations of capitalism (i.e., between SMEs and LMEs). The finding that the effect of a partisan composition of government varies among sociopolitical regimes (it matters in LMEs but not in SMEs) is also important in understanding the working of the median voter theorem, as specified in the previous section.

In a subsequent study Pontusson et al. (2002) also found that higher levels of unionization and wage bargaining and larger shares of public sector employment reach their equalizing effects primarily by improving the relative position of unskilled workers (who constitute the lower tail of the distribution), but partisanship (most notably the participation of the Left in government) has an equalizing effect on the upper end of the distribution by constraining the wage growth of the highly skilled. In centralized wage bargaining systems the Left governments seem to be successful in controlling changes at both the upper (taxation, etc.) and the lower (minimum wages, etc.) tails of the wage distribution.

Reflecting the fact that parties traditionally considered "Left-wing" became increasingly heterogeneous in their ideological beliefs and policies throughout the last decades, Tepe and Vanhuysse (2013) reclassify them by reweighting their nominal positions with their ideological stances/declarations in their party manifestos (data taken from the Comparative Manifesto Project). Also, the same authors aimed to identify strategies of Leftist parties and of trade unions with regard to their effect on EPL (assumed to favor insiders) and active labor market policies (ALMP; assumed to favor outsiders). Analyzing data from a sample of 20 OECD countries between 1986 and 2005, they found (in line with Rueda, 2008) that the Left party power variable has no effect on outsider-favoring ALMP spending in general and a negative effect on job creation programs (which contradicts what PRT theorists suggest). However, as they emphasize, larger and more strike-prone unions tend to increase ALMP spending overall, specifically in those dimensions that help their members: employment assistance and labor market training (Tepe and Vanhuysse, 2013).

19.5.5 Redistribution Via Taxes and Transfers: Technical and Efficiency Aspects

The question of why and in what direction redistribution changes the pre-tax and pre-transfer income distribution depends largely on the interplay of various political forces that are able to influence the political process. The question of how and with what effectiveness it happens is more of a technical nature. This section describes some aspects of effectiveness, many of which are not straightforward right from the outset.

The identification and measurement of redistribution presupposes a counterfactual that exists before the redistributive action of transferring money from taxpayers to benefit recipients takes effect. However, the pretransfer distribution already is influenced by regulatory acts (relating to interhousehold transfers such as alimony and others such payments, to employer–employee relationships such as regulations of wages or working conditions, to supply and demand in various markets such as rent control in housing markets, etc.), the operation of which contributes to the shape taking place before conventionally defined income distribution starts to be measured.[66] Further, the features of "pre-redistribution" are embedded into a broader context such as informal norms of responsibility over the welfare of others (younger or older family or local community members, the poor or the handicapped, etc.); the actual role of such forms of informal solidarity varies across countries. These caveats need to be mentioned at the outset, although no extensive coverage can be given to them in what follows.

Broad forms of redistribution (and of welfare states) can be classified into two categories: the "piggy banks" and the "Robin Hoods" (Barr, 2001). The piggy bank approach puts the focus on smoothing consumption and on insurance against risks prevalent in various stages of the life cycle. In its ideal form it has an effect on life cycle distribution of incomes but does not lead to interpersonal redistribution. The other type (the Robin Hood approach) focuses is on redistribution between various social strata (most commonly from the rich to the poor).

Our image (and, even more, our evaluation) of the extent of redistribution is greatly affected by the perspective from which we see incomes and benefits. Consider the largest item—pensions—as an example. In actuarially fair pension insurance systems there is no interpersonal redistribution involved. Under given parametric regimes of accrual rates, retirement ages, compensation rates, etc., people save for income security during their old age. But putting this income transfer into a cross-sectional frame produces a false impression of the extent of redistribution between richer and poorer segments of the society at a given point in time. In the same vein, the perspective has to be clear when evaluating the redistributive role of sickness insurance, education finance (especially at a higher level), and many other fields.

Furthermore, for cross-country comparisons of income distribution, it should be made clear that countries differ in the mix of the characteristics described above (systems such as the Danish tax-financed welfare states are more the Robin Hood type, whereas Bismarckian systems and to a lesser extent the Beveridgean systems are more piggy bank types), although no really ideal types exist. However, changing the perspective also changes our images of the redistributive effects of the various welfare state arrangements. (See Whiteford, 2008 for more on this.) The extent to which welfare states focus on redistribution among versus between people in a lifetime perspective varies considerably (roughly half in Australia but two-thirds in the United Kingdom and four-fifths in Sweden, taken from a lifetime perspective; see Hills, 2004; Ståhlberg, 2007). This also hints to what extent we can expect welfare states to modify income distribution in a long-term perspective.

[66] This "counterfactual problem" in welfare state research has been discussed by Bergh (2005) and Esping-Andersen and Myles (2009). See also Lambert et al. (2010) and Förster and Mira d'Ercole (2012).

For explaining the distribution of current incomes (our focus in this chapter) it is mostly the Robin Hood–type welfare state activity that matters.[67] Among the many related issues (mostly treated in Chapters 23 and 24 on antipoverty policies and microsimulation, respectively), our focus remains on the effect of redistribution on incomes. We focus on the following questions.

– What overall first-order effect does redistribution have on (initial, cross-sectional, "virgin") income distribution?
– What feedback/secondary effects of redistribution can be identified?

To measure redistribution, setting up a proper income accounting framework is crucial. The commonly used framework (see OECD, 2008, for example, but earlier in Atkinson, 1975) starts from (1) factor incomes (i.e., gross wages, salaries, self-employment and property incomes, adding private occupational pensions to arrive at (2) market incomes, which are supplemented by social benefits, private transfers and miscellaneous cash incomes, resulting in (3) gross income, from which the deduction of various taxes (on wages and/or incomes, by employees and/or employers) results in (4) disposable cash incomes (see Förster and Whiteford, 2009 for more on this framework). Attempts to measure redistribution compare various elements of the above to assess the immediate (direct, first-order) effects of redistribution.[68]

19.5.5.1 Overall, First-Order Effects of Redistribution

After comparing pre-redistribution (market income) inequality to post-redistribution (net disposable income) inequality, Whiteford (2008) concluded that redistribution reduces inequality by roughly one-third of the "original" inequality (ranging between 45% in Denmark, Sweden and Belgium and some 8% in Korea [Whiteford, 2008]). These results refer to the entire population and thus include the effect of public pension transfers, which, as argued earlier, blurs the picture. The OECD (2011, 2013) showed that the redistributive effect of public transfers and taxes for the working-age population—thereby excluding public pensions to a large extent—amounted to, on average, little over a quarter across OECD countries in the late 2000s, reaching close to 40% in some Nordic and continental European countries.

Immervoll and Richardson (2011) showed that redistribution (as measured by the difference between Gini coefficients before and after redistributive measures, whichever is appropriate) increased between the 1980s and the mid-2000s in general across the OECD. However, the pace of increase of market income inequality to a large extent

[67] Mostly, but not exclusively. Life cycle income smoothing mechanisms also have cross-sectional income distribution effects. Consider the immediate effect of pensions on the relative position of the elderly. However, social insurance instruments are better judged by their own standards: their ability to smooth consumption over the life cycle.

[68] Most empirical studies are, however, confined to the effect of cash transfers and direct income taxes. Publicly provided services (in-kind transfers) also play an important redistributive role. While the inequality reducing effect in general is lower than that of cash benefits, it is still sizeable and amounts, for instance, to on average 20% of OECD countries in the 2000s (see, e.g., OECD, 2011; Förster and Verbist, 2012).

exceeded the increase of redistribution during the period. Especially during the periods between the mid-1990s and mid-2000s, the redistributive strength of tax benefit systems decreased in many countries (in the latter period the weakening redistribution contributing to inequality increased more than market income inequality increased in itself).

Regarding the redistributive effectiveness of the two sides (taxes on the one hand and expenditures on the other), the OECD (2008) and Whiteford (2008) found redistribution achieved by public cash transfers was twice as large as redistribution achieved by income taxes (except, among the whole OECD country range, the case of the United States, where taxes play a greater role). Immervoll and Richardson (2011) found that the effect of benefits on inequality was much stronger than social contributions or income taxes,[69] despite the fact that taxes and contributions were larger compared with household incomes.[70] Partly relating to this, the overall effect of the tax/benefit system on the various parts of the income distribution was found to be more prevalent in the bottom tail than in the top of the income distribution (Immervoll and Richardson, 2011).

Nevertheless, Fuest et al. (2009) highlighted that the differential effect of taxes and transfers on redistributive outcomes is sensitive to the methods applied. In their study of 25 EU countries on the basis of the 2007 wave of the EU-SILC survey, their analysis, following the traditional redistribution accounting framework (see Förster and Whiteford, 2009), confirms that benefits are the most important inequality-reducing factors. However, when applying factor decompositions described by Shorrocks (1982) (i.e., when determining what roles various factor components play in determining overall inequality), they concluded that benefits play a minor role (if any) in redistribution. This later procedure results in a much larger role for taxes and contributions in inequality reduction in almost all countries (Fuest et al., 2009). Among the explanations, they argue that while in a traditional accounting framework an equally distributed social transfer tends to have a positive effect on final inequality, to achieve a redistributive effect in a decomposition framework requires a definite negative correlation of transfers with incomes. There has, however, been criticism with regard to policy interpretation of results based on the decomposition framework, which estimates the contribution of equally distributed income sources to overall inequality, by definition, as zero. This is regarded as not being intuitive because a flat-rate benefit that is "added" to unequally distributed pre-transfer income would normally be expected to decrease inequality.[71]

[69] Similarly, Mahler (2010) also found a much smaller redistributive effect for taxes than for social transfers.

[70] The corresponding effective tax rate is measured by dividing all taxes paid by all pretax income (of households, for both items). The analysis by Immervoll and Richardson (2011) takes into account the country-specific interactions of taxes with benefits and legal differences in sequencing, for example, the fact that some benefits are taxable while others are not.

[71] Another point is that the results by Fuest et al. (2009) are based on the coefficient of variation, which is highly sensitive to outliers at the top, and this mere fact can lead to somewhat misleading interpretations. In addition, the fact that in certain countries the EU-SILC is based on registers that better capture top incomes, the direct cross-country comparison of redistributive effects of benefits estimated by effects on a tail-sensitive measure can be another reason for caution.

Based on LIS data comparisons, Lambert et al. (2010) suggested at the outset that empirical literature on the relationship between income inequality and redistribution is inconclusive. Given the fact that pre-redistribution (i.e., pre-tax and pre-transfer) income inequality can, by definition, be counterfactual only, they suggest a method called "transplant and compare" for measuring the "true" effect of redistribution, independent of the starting level of inequality of the observed countries. When income tax systems are evaluated according to their own pre-tax/-transfer inequality baseline, redistributive effects of personal income taxes seem to be stronger in more unequal countries for most of the measures they applied. When harmonizing the baselines across countries, they found a weaker relationship.

Based on an analysis of an unbalanced panel of 43 upper-middle- and high- income countries for the period 1972–2006, Muinelo et al. (2011) put the issue of redistribution and inequality into a broader context. After estimating structural equations to model the role of fiscal policies in economic growth and inequality, they found that increasing the size of the public sector (defined as direct taxes and expenditures), while decreasing inequality, harms growth. However, the effect of indirect taxes on both growth and inequality was found to be insignificant. Public investment of general government as a share of GDP, however, is shown to have an equalizing effect without harming economic growth. For a more restricted data set (an unbalanced panel of 21 high-income OECD countries for the period 1972–2006) and with a different variables structure for fiscal policies, Muinelo et al. (2013) found a positive correlation between lower levels of inequality and the size of the public sector (defined in terms of expenditures and taxes per the GDP). They also found that an increase of distributive expenditures (public spending on social protection, health, housing and education) to reduce income inequality in high-income welfare states had no a clear harmful effect on growth. At the same time, they found that an increase in nondistributive expenditure (general public services, defence, public order, economic services) decreases economic growth while increasing income inequality, irrespective of the financing sources (direct or indirect taxes) of expenditures.

Afonso et al. (2010) attempted to estimate how effectiveness (success in achieving program objectives) and efficiency (the degree to which the use of available resources maximize their objectives) of public spending programs is achieved in various countries. According to their propositions, higher social spending is associated with a more equal distribution of incomes across the OECD countries. Southern countries are shown to perform less well in terms of efficiency than Nordic countries. For the Anglo-Saxon countries, output efficiency (the degree to which outputs can be maximized with given inputs) tends to be low, whereas input efficiency (the degree to which a given output can be maintained with decreasing inputs) tends to be high.

On the basis of an analysis of 25 OECD countries, Goudswaard and Caminada (2010) found that total public social expenditures have a strong positive effect on redistribution (and inequality reduction). At the same time, countries with higher private social expenditures have lower levels of redistribution. When excluding services (health expenditures in their analysis), social expenditures (public and private) were shown to make a somewhat smaller contribution to inequality reduction. However, the effect of spending on services did not seem to have a strong effect on their results. The various elements of social expenditures have different contributions; public pensions have larger effects and unemployment benefits and labor market programs have smaller but still positive effects. The sign for private pensions was shown to be positive, implying an inequality-increasing effect.

19.5.5.2 Back to Politics: The Paradox of Redistribution

With regard to the effect of welfare spending on poverty and income distribution, an influential article by Korpi and Palme (1998) pointed out an apparent paradox: they found that targeted benefit systems may have achieved less redistribution than more universal ones, based on available data for the 1980s. Kenworthy (2011) confirmed this finding for the original 10 OECD countries Korpi and Palme analyzed for the 1985–1990 time span. However, Kentworthy showed that that this inverse relationship between targeting and redistribution has weakened by the mid-1990s and then disappeared by 2000–2005. With refinements of the measures, extensions of the country coverage and robust checks of sensitivity to alternative income definitions, Marx et al. (2013) argued that the claimed empirical relationship as such no longer holds. On the methodological side they indicated that the outcomes are not only sensitive to operationalization (i.e., definitions of the counterfactual) and data sources (such as differences between LIS and EU-SILC data) but also to the country selection (inclusion of southern and eastern European countries reveals patterns that are different from each other and also from the previously involved country groupings). On the policy side, they argued that the nature and effects of targeted programs also substantially changed as the decades elapsed (with more emphasis on incentives and changed focus targeting in-work groups started to enjoy more support from middle class electorates as well). With better data, more refined analytics and broader coverage, Marx et al. argued that it is the differential efficiency of various targeted programs and of different country experiences that has to be explained in future research.

Identifying and measuring inequality-reducing effects of redistribution may become prohibitively difficult in the frame of understanding of welfare regimes (Esping-Andersen and Myles, 2009). A full analysis should involve an analysis of taxes and transfer schemes and services, all analyzed simultaneously in a complex setting where state activities are embedded into general societal functioning, producing welfare outcomes jointly with the market and the family. Under these circumstances, the same egalitarian commitments of two different states may produce different results (Esping-Andersen and Myles, 2009). This makes systematic accounts very difficult, calling rather for analysis in a case study fashion. It is therefore important to understand the nature and operation of welfare state interventions at a program level before generalizing to the level of welfare regimes.

19.5.5.3 Second-Order Effects of Redistribution: Labor Market Responses

The above findings may, however, misguide us in the understanding of redistribution if we do not pay attention to the fact that there are second-order effects that also have to be specified and analyzed. The immediate effects (as above) are "overnight" hypothetical gains to recipients (say, of social assistance) and costs to contributors (say, taxpayers). Groups on both sides may vary (according to what type of redistribution is at stake). However, redistribution can also induce second-order effects as actors when noticing changes in costs and benefits their actions will adopt (rich people may change the way they receive their incomes to lower their effective tax rates, whereas poor people might change their labor supply, etc.). Regarding second-order effects, there are many assumptions and fewer tests (except, perhaps, tests of the Laffer curve, assuming high elasticity of labor supply to changes in marginal tax rates).

When modelling second-order effects, Doerrenberg and Peichl (2012) found no significance for the progressivity of income taxes, concluding that, for tax variables, the second-order (behavioral) effects might be larger than they are for expenditures. Niehues (2010) concluded that increased specific targeting of low-income groups is not associated with lower postgovernment levels of inequality. From this, her indirect conclusion is that there might be second-order (potential disincentive) effects in the case of means-tested benefits. However, her analysis of the overall effect of social transfers shows strong equalizing effects that largely outweigh second-order effects.

Blundell (1995; and Blundell et al., 2011) examined potential effects of income taxation on labor supply (extensive margin [decisions to enter labor market from the outside] and intensive margin [work effort decisions of those already in the labor market]). They found that labor supply elasticities for women at both margins are larger than elasticities for men. The overview by Blundell (1995) lists a number of factors why individual labor supply responses to changes in marginal tax rates is very complex (fixed costs of work, life cycles aspects of savings, demographics and wealth accumulation, on–the-job human capital and seniority, the role of unions and collective bargaining, as well as benefit usage and effective tax rates). All these elements characterize the actual operation of the redistribution, making generalized judgements of the secondary effects of redistribution almost impossible. It is even more difficult to draw any further conclusions with respect to inequality effects, given the large number of corresponding assumptions in addition to the above (the interplay of behaviors/demographics and of the labor market effects and income effects, etc.).

Starting from the assumptions that labor supply elasticity is higher at the bottom than at the top and that higher redistribution may shift employers away from social responsibility, Doerrenberg and Peichl (2012) expect negative second-round effects of redistribution on inequality, that is, increasing inequality. However, in an unspecified panel of

OECD countries for the period of 1981–2005, they found that redistributive policies' first-order effects (we might call it "overnight incidence") remain dominant when taking into account the offsetting second-order effects (i.e., behavioral repercussions). They concluded that a 1% increase in public social spending reduces inequality in the order of 0.3% in magnitude overall. Care must be taken when interpreting the magnitude of second-order effects when they are attempted to be put into a conventional redistribution framework. Consider for example the case when market income inequality is contrasted with disposable income inequalities. The differences of the Gini coefficients calculated for these two elements may already entail behavioral reactions from the past and they may also provoke reactions in the future. Therefore, introducing the time dimension is important, especially for the understanding of the second-order effects.

19.5.6 Structural Societal Changes

There are a number of reasons why changes in social structure have direct (via changing composition and the changing relative sizes of various societal subgroups) or indirect (via changing behaviors) effects on income distribution. Below is a list of examples of both direct and indirect effects, in the order of the demographic groups in question.

In ageing societies, depending on the concrete institutional arrangements of the pension systems, the growth of the elderly population may contribute to lower aggregate income inequality, given the fact that in most pension systems the inequality between pensioners is smaller than inequality among the active-age population, but it may also contribute to higher inequality because pensioners, on average, have lower relative incomes. Also, the growing imbalance between social insurance recipients and social insurance contributors (or taxpayers) induces shifts in retirement ages—a fact that also has a direct consequence on pensions-to-wages ratios and, through this, on income distribution. Furthermore, the shifting of the age balance of the electorate affects the political power of the elderly who, in elections, may have a stronger voice on public expenditure preferences; this points towards the direction of the relatively better situation of the elderly compared with the income situation of the younger generations.

Another example is that changes in family structures can also have direct and indirect effects. The long-term trend of the breakup process of traditional large families results in a larger number of societal units with a smaller average size. The unit of analysis for income inequality (as opposed to wage inequality) is the household. The changing household structure in a country (decline in household size, breakup of traditional family forms such as the breadwinner model, etc.) affects the unit of measurement, and this may have an immediate effect on household inequality, even if there is no change at all in wage distribution. The same holds for changes in household composition by labor market attachment; for example, an expansion of female participation in the labor force, depending on the distribution of it, will itself alter distribution. In addition, and parallel to the breakup of larger units, an additional strain on the welfare state may arise, given the duties of modern states in taking care of vulnerable citizens (should the breakup take the form of the increase of single-parent families and/or the share of elderly single households).

Further, a general education expansion (which was massive in the past 50 years in the OECD area) not only changes the structure of subgroups with higher and lower skills but also contributes to deeper societal trends: more educated voters might become more interested in politics, with stronger opinions on economic or social policies, etc. Related to this, the emergence of a broader or shrinking middle class not only has a measurement consequence but the middle class change might also induce behavioral and attitudinal consequences.

Finally, the change of the composition of the population by origin of birth as a result of international migration can lead to income distribution changes, depending, of course, on which parts of the income distribution of the recipient country the migrants enter. Also, changes in the attitudes or ethnic composition of societies might urge politicians to reflect these attitudes in changes in their policies.

While there are a large number of studies of some particular aspects of these trends, relatively few systematic accounts of the effects of social structures in income distributions are available. When assessing the role of population structure changes on summary measures of inequality, the OECD (2008) emphasizes that income inequality exists between and within demographic groups (of various ages or by sex, for example). That study presented simulation results, considering population demography as "frozen" at the start of the observation period (mid-1980s or mid-1990s, depending on the country) to show the independent effect of changing population composition on income inequality. This highlights that changes in demography (ageing and household structure change combined) contributed to higher income inequality in most countries. It also showed that the effect of the change of household structure seems to be larger than the effects of ageing. Changes in population structure were driven by the increase of single-parent households, a key trend in determining the overall demographic effects.

The effect of demographic trends on income inequality has been studied by a number of papers in the past two decades (see Burtless, 2009 and OECD, 2011 for an overview), but the number of systematic cross-national accounts is small. It has been shown for the United States (see Karoly and Burtless, 1995; Burtless, 1999) that the increase in the share of single households was an important contributor to the increase of inequality. Similar trends were shown for Germany (Peichl et al., 2010) and Canada (Lu et al., 2011), although the latter was not confirmed by another study of five OECD countries (including Canada) by Jantti (1997).

Marital sorting or "assortative mating," that is, the growing tendency that people are married to spouses with similar earning levels, can also contribute to higher inequality, which has been documented in a number of country-specific studies. Schwartz (2010), for instance, found that, for the United States, assortative mating contributed one-quarter to one-third to higher earnings inequality among married couples, with the main contribution occurring at the top of the distribution. A review of some other country-specific articles by the OECD (2011) lists a number of studies showing that an increased similarity of spouses' earnings in households contributes to widened inequality (OECD, 2011) Cross-country evidence, however, is rare. The role of assortative mating can be illustrated by counterfactual simulations (Burtless, 2009; Chen et al., 2013b). As these simulations show, assortative mating may have nontrivial effects on inequality. The OECD (2011) provides an overview of the literature, which indicates that a number of studies show that increased resemblance of spouses' earnings had an inequality-increasing effect, although there is a wide range of estimates as to the relative weight of this effect.

OECD (2011, chapter 5) looks into this issue from a broader perspective, analyzing the transmission of earnings inequality from individuals to households in 23 countries. Results drawn from primary-order decompositions show that labor market factors outweigh demographic factors for determining increased household earnings inequality by far; the major driver behind household earnings inequality is the increase of male wage dispersion (this contributes one-third to one-half to the overall increase of household earnings inequality). A second major factor, but one that works in the opposite direction, is the increase in women's employment in most of the countries under scrutiny. This had an off-setting, that is, equalizing, effect everywhere. Finally, demographic factors also are shown to contribute to inequality. Both the effects of the more widespread assortative mating and the change of household structure played a role, directing towards a larger inequality, though this effect was assessed (OECD, 2011) to be much more modest than labor market–related changes.[72]

In their recent article, Greenwood et al. (2014) concluded that assortative mating increased between 1960 and 2005 in the United States, with an increasing effect on inequality; comparing inequality figures based on assortative mating with inequality figures based on random matching, the estimated difference increased considerably, implying that part of the inequality increase in the United States can be accounted for by increased marital sorting.

In his LIS-based analysis of 18 rich (mostly OECD) countries, Brady (2006) tested the effect of various structural factors on the lower tail of the income distribution. He found that an increase in employment in general, and female employment in particular, reduces income poverty. After controlling for institutional factors (welfare state variables) and economic factors, this was found to be the largest single item with the largest poverty-reducing impact. On the other hand, the growth in the share of the elderly population and the increase in the share of children in single-mother families had an effect on increasing the poverty headcount. When concluding, however, he stressed that the welfare state has a larger effect than structural factors.

[72] The effects of assortative mating and other household structure changes taken together are estimated to count roughly half as much as the effect of increased male wage dispersion alone.

The equalizing effect of women's participation in employment also is documented in other recent cross-country studies. On the basis of a counterfactual analysis of 20 OECD countries, Chen et al. (2014) found that if female labor force participation had not increased in the past 20 years, household income inequality would have increased by 1 point more on average than it actually did.

Esping-Andersen (2009) pointed to the importance of demographic shifts in society, sometimes even counterbalancing the effects of large trends such as globalization and technology. The changing role of women in terms of increased labor market participation, domestic work, marriage and education has a large role in the formation of inequalities. As he argues, the process, characterized by women's commitment to longer work careers and to their increased participation in (higher) education, via more equal division of domestic work between spouses and a greater degree of assortative mating, leads to a lower level of inequality within the family (i.e., among men and women), but it also leads to higher level of overall inequality in the society. The latter trend is induced primarily by the fact that it is the higher-educated and higher-income women among whom the process runs first, leading to widening inequalities between women with higher and lower social status. From this it follows that observed cross-country differentials in income inequality also reflect the state of what he terms the "incomplete revolution" of changing gender roles (Esping-Andersen, 2009). A next step in this reasoning could be that because societies differ according to their dominant family patterns (the two extremes being the male breadwinner model/nuclear family on the one hand and a model characterized by dual earner models and shared domestic work on the other), so too do their inequality patterns differ. This conclusion remains to be proven by further empirical comparisons.

The effect of demographic and household formation changes in households have, in turn, different consequences for inequality and income dynamics, depending on the differential institutional structures in various countries. As DiPrete and McManus (2000) concluded in their US–Germany comparisons, the chances of individuals and households responding to "trigger events" (such as partner losses, unemployment, etc.) are different in institutional settings relying more on the market than in countries having more elaborate welfare arrangements. The effect of shifts in income and material well-being, triggered by household employment and household composition changes, is mediated by tax/transfer schemes as well as by private responses to these events. As DiPrete and McManus highlight, the relative role of labor market events, family change and welfare state policies in income dynamics also depends on gender.

The effect of migration on inequality in donor and in recipient countries depends on the skill composition of migrants and native populations, on the process and speed of integration of migrants into the host labor markets, on differential household composition of migrants and of natives, among other factors. Also, the balance of inward and outward migration and the institutional structure is of major importance. Not only the share but also the skill composition of migrants varies substantially across countries. This makes drawing general conclusions on the effect of migration on income distribution very difficult (if not impossible). The effect—if it exists—is thus very much country and context dependent. The vast empirical migration analysis literature focuses on these elements on various target variables such as labor market outcomes, poverty and tax/benefit systems, but they very rarely have the ambition of modelling the full impact of migration on overall income inequality (Chen, 2013).

A few models, however, are formulated to reach some broad general conclusions. Kahanec and Zimmermann (2009) introduced a model with heterogeneous labor markets. Their prediction is that highly skilled immigration can contribute to a decrease in inequality in the receiving countries. The argument (although with many caveats about complementarities between skilled and unskilled labor and about institutional and social histories of the various country contexts) stresses that, in OECD countries where skilled labor is abundant, the degree of the labor market assimilation of immigrants into the host country is key in determining the true long-term effect of migration on inequality. There is a much less general conclusion that can be offered for unskilled migration. Kahanec and Zimmermann (2009) concluded that the effects can be expected to be ambiguous.

As a conclusion of a thorough literature review, Chen (2013) identified a number of challenges for the assessment of the effect of migration on inequality. As he concludes, most assessments are partial (focus on relative wages rather than on the full distribution) and mostly cross sectional (and, as such, overlook the earnings potential and lifetime earnings of migrants). The review suggests building integrated micro-/macrosimulation models to assess the full effects of migration on income inequality.

19.6. CONCLUSIONS: MAJOR FINDINGS FROM THE LITERATURE SURVEY AND IMPLICATIONS FOR FURTHER RESEARCH

19.6.1 A Summary of Findings and Propositions from the Overview of Studies Providing Multicausal Explanations

This section summarizes the main findings presented above from the most important recent studies that provide multicausal explanations and provides a combined analysis of the relative weights of the various arguments set out in Section 19.5. For the purpose of the summary, we differentiate between three levels of explanatory factors. On the first, broadest level (represented by the diamonds in Figure 19.1), there are six different groups of factors:

1. structural macroeconomic sectoral changes
2. globalization and technology change
3. labor market and other relevant institutions
4. politics and political processes
5. tax/transfer schemes
6. demographic and other microstructural changes

As indicated in Section 19.1, we may think of the above factors as "underlying" causes of inequality change. On the second level, there are elements within each of the six broad groups (such as FDI, technology, trade, etc., for globalization or such as unionization, unemployment benefits, employment protection legislation, etc., for labor market institutions). This second group could be included under the umbrella of "proximate causes" of inequality or "hints" at causes.[73] Finally, there is a third level, on which the various authors operationalize their models, that is, where they chose the appropriate variables for their models, which are, in most cases, necessarily second-best proxies of the second-level factors. In what follows, we summarize the results of the level of abstraction represented by the first level. While doing that, we also report findings for the interactions between the effects of the various variable groupings as far as they are available.

As for the major hypothesis of structural macroeconomic sectoral changes (i.e., sector bias and sector dualism, as proposed by Kuznets), the evidence is inconclusive. A large part of the literature (half of 30 studies reviewed by Atkinson and Brandolini, 2009 and 19 studies in Hellier and Lambrecht, 2012) tests the Kuznets hypothesis, but sector dualism does not seem to find support.[74] Alternative explanations of the great U-turn therefore have been investigated in various articles in the past 15 years. The most influential hypotheses of these alternatives related the reversal of inequality trends to developments of globalization and of trends in skill-biased technology change to changes of (labor market) regulations and institutions.

As for the debate on globalization versus technology, there has been a move away from trade-focused explanations to technology explanations during the 1990s. In the 2000s, several authors changed track from their earlier views that the effect of trade on inequality was modest at best (Krugman, 2007; Scheve and Slaughter, 2007). They now suggest that trade-induced phenomena such as outsourcing may have had a more significant effect on income distribution than formerly assumed. That said, while under the pure aspect of trade costs, off-shoring all tasks that are technically off-shorable may indeed be possible, this will not always make sense from a business point of view, especially when transaction costs and economies of scope are taken into account; the assumed effect of a surge in off-shoring may therefore be exaggerated, as argued by Lanz et al. (2011).

At the same time, technological change now is more often understood as endogenous and interacting with trade. More generally, the key issue today is no longer identifying which trade or technological change was the main culprit in increasing inequality, but rather to identify the channels through which these two operate and interact in their effect on inequality (see Chusseau et al., 2008).

[73] These notations follow Cornia (2012).

[74] However, Nollmann (2006) and Rohrbach (2009) propose a focus on knowledge sector dualism and bias.

The effect of *education*—human capital accumulation—on inequality is not linear and, because of different composition and wage premium effects at different times, can first be disequalizing and then equalizing, analogous with the Kuznets process. That said, none of the studies covering the set of OECD/EU countries suggest a disequalizing role for the growth in average educational attainment over the past three decades; on the contrary, in their majority they propose a rather equalizing role. Human capital can be seen as a complement to technology. Increases in human capital and in the supply of skills are necessary to decrease and eventually reverse the pressure to higher inequality that stems from technological change.[75]

While it is widely recognized that institutions matter, the weight attached to this factor in econometric studies has long been limited. A majority of (but not all) studies finds significant negative associations, in particular with wage inequality, through direct or indirect effects of union density/coverage, wage coordination/centralization and EPL. Checchi and Garcia-Penalosa (2005) and the OECD (2011) found the weakening of employment protection and the decline in unionization increased wage dispersion, mostly having effects at the lower ends of the distribution of wages. It has, however, also been emphasized that when observed in a broader context (i.e., concentrating on combined employment and dispersion effects of institutional changes), the results were inconclusive because employment and inequality effects of institutional change tended to net each other out (OECD, 2011). Also, Checchi and Garcia-Penalosa (2008) suggested that the combined effects of institutions on factor income inequality are weak, whereas the income distribution effects of high tax wedges (which could be expected to serve larger redistribution to favor lower segments of the labor markets) also has controversial effects (high-wage workers are able to pass on tax burden to their employers, while the overall tax wedge effects can contain considerable unemployment increases).

All in all, it is shown that for inequality trends, developments in political processes are of key importance. How preferences of the electorate are recognized, processed and translated into policies (which, in turn, shape labor market and welfare state institutions) do play an important role in redistributive institutions and, ultimately, in inequalities. Indirect proof of this is found in the fact that many tests trying to find a direct relationship between initial and post-redistribution inequalities have been shown to be inconclusive. While some of these failures can be explained by problems of specification, of identification of the various factors or of data, there are a number of substantive elements of the political system that may have a special role in defining inequalities. Among these, the

[75] It can be suspected, however, that this is conditional on the stage of the "race between education and technology" change (Tinbergen, 1975). Most of the studies reviewed here refer to the OECD area for the 1980–2008 period, a rather fortunate period and set of countries where higher education expansion was to a great extent capable of keeping pace with the upwards pressure of the technology revolution. In different countries and in different periods, the results of this race may be less positive for inequality outcomes.

differential mobilization of voters from various parts of the income scale seems to be of a crucial importance (Pontusson and Rueda, 2010; Mahler, 2008). Also, how the actors of the political arena perceive their core constituencies is important. If the parties from the political Left perceive the mobilization of the poor on the ballots worth going for, they may put the issue of redistribution to the poor at the center of their political agenda.

The identification of the Left and the Right may easily turn out to be problematic, especially when representation of the various labor market segments is taken into account (Rueda, 2008). Given the fact that parties sometimes pick up interests of insiders (such as active earners) as opposed to the interests of outsiders (such as the inactive earners and the unemployed), redistributive outcomes might come about as results of sometimes contradicting tendencies of redistribution from the rich to the poor and of legislation to support the interest of the insiders of the labor markets.

When analyzing actual redistribution processes, the definitions of the pre- and post-redistribution inequality (in other words, the accounting framework in which the redistribution processes are understood and interpreted) has been identified as crucial to the measurement of the effects of redistribution (Whiteford, 2008; Immervoll and Richardson, 2011; Kenworthy and Pontusson, 2005). It also has been emphasized that redistribution might have a number of second-order effects. The results of redistribution analyzes have shown that redistribution reduces inequality overall in all OECD countries, although to a varying extent, depending on concrete institutional settings. It was found that "original" inequality (if it exists at all) is reduced by an order of magnitude of some one-third by redistribution (ranging between 45% in some northern and continental European countries to ~8% in Korea; see Whiteford, 2008; OECD, 2011).

The redistributive effectiveness of the two sides (taxes and benefits) has been shown to be different: cash transfers (in all countries but the United States) are estimated to have much larger first-order effects on inequality than taxes (Whiteford, 2008; Immervoll and Richardson, 2011).[76] Among public social transfers, public pension programmes achieve the largest redistribution; however, the interpretation and evaluation of these differs and is dependent on the chosen perspective of Robin Hood or piggy bank welfare states.

There are second-order effects of redistribution, such as those resulting from behavioral adjustment on the contributor side (taxpayers) or the recipient side (social assistance beneficiaries). Some studies are able to show the existence of second-order responses, the magnitude of which, however, seems to be relatively small (Doerrenberg and Peichl, 2012). The measured effects of taxation on labor supply (which is clearly an important area of potential behavioral repercussions) imply that social embeddedness of institutions is noticeable. Studies by Blundell et al. (2011)

[76] This is also confirmed by other studies (Mahler, 2010; Goudswaard and Caminada, 2010). The latter study also shows that countries relying mostly on public social expenditures achieve higher levels of redistribution than countries relying more on private social transfers.

highlighted that behavioral elasticities for women are larger with regard to both decisions about entering the labor markets (extensive margin) and changing work efforts on the labor markets (intensive margins).

An important aspect in redistribution research is how the change in size and techniques of tax transfer schemes have contributed to changes in overall inequality. As highlighted by the OECD (2011), changes in redistribution can be seen as causal factors for increasing inequality during the period before the breakout of the economic recession in 2008. The redistributive power of the welfare state was weakened in the period between the mid-1990s to mid-2000s. While in the period between mid-1980s and mid-1990s the share of increased market income inequality offset by taxes and transfers was measured at a level of almost 60%, this share declined to around 20% by the mid-2000s (OECD, 2011).

The social context can also be captured by the effects of changing demographic composition (by age, household types, etc.) and of changing demographic behavior (household formation, assortative mating, etc.) on inequality. While the (composition) effects of ageing and of household composition are estimated to have an inequality-increasing effect (Lu et al., 2011; OECD, 2011; Peichl et al., 2010), the results of some of the discussed behavioral trends (assortative mating) are less clear-cut, but in general also are shown to have an effect on inequality change, mostly as disequalizing effects. Some scholars present the results of the "incomplete revolution" of women's changing role in labor markets and in families as equalizing *within* the households (because of more equal divisions of domestic labor) but disequalizing *among* households (because of differential behavioral reactions of women with higher and lower status [Esping-Andersen, 2009]). Taken together, when modelling the inequality effects of changes in demographic composition and behavior on the one hand and labor market related changes on the other, the OECD (2011) concludes that the former seems to explain much less of the increase in inequality than the latter.

In a nutshell, this is what we found at the first level of factors identified at the beginning of this section (and in the diamonds of Figure 19.1). To give a brief summary assessment of the results found in the studies published over the past 10–15 years, Figure 19.3 provides an idea of the direction of causal factors of inequality that were identified. This summary remains qualitative and cannot be based on quantitative assessment because the multitude of studies use various and different methodologies, estimation methods and data, as well as varying country coverage. Further, it is in part our own subjective assessment. As a convention, positive/negative association means disequalizing/equalizing. "Significance" has to be understood here (and elsewhere in the text) as a statistically significant association, notwithstanding the relative size of a coefficient. "Inconclusive" means that roughly as many studies report (significantly) positive as negative effects. Further, this assessment is based as much as possible on studies covering the restricted sample of OECD/EU countries.

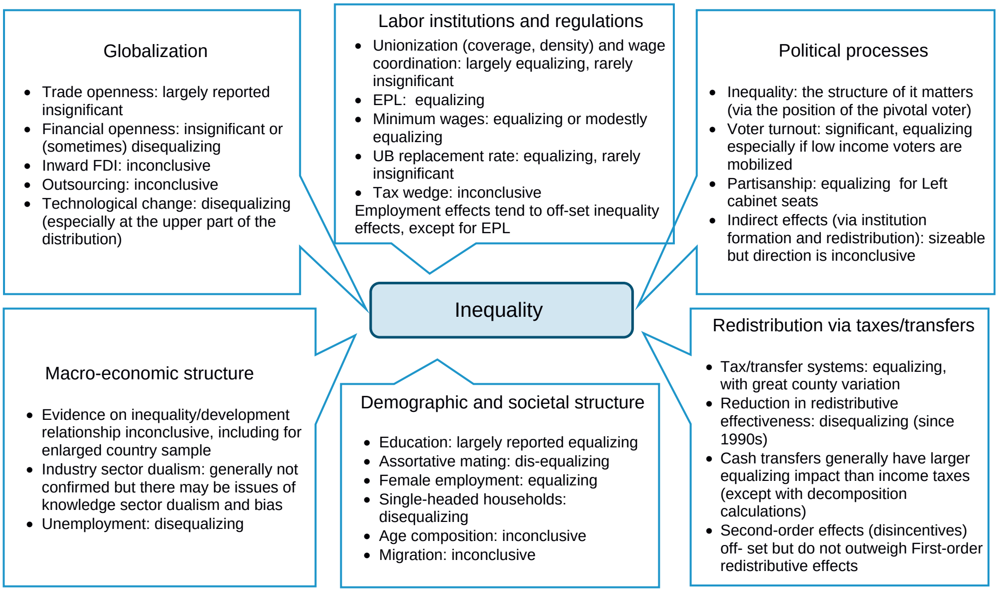


Figure 19.3 Drivers of inequality: a qualitative summary of results for OECD countries reported in recent studies. EPL, employment protection legislation; FDI, foreign direct investment; UB, unemployment benefit.

A first glance at Figure 19.3 reveals that inconclusiveness prevails for many possible drivers of inequality, that is, the large number of recent empirical, cross-country studies report contradicting results, which can often but not always be traced back to different country samples, time periods, data and methodological specifications. In particular, for those factors for which there are more complete and fairly direct measures at hand (such as measures of trade openness or financial openness), there is little clear effect reported, whereas for factors where more proxy-type measures need to be used (such as technology), there seem to be more significant findings. One is tempted to detect some sort of Heisenberg principle: the sharper we can measure a variable, the less effect will be found.

As mentioned above, the summary assessment in Figure 19.3 refers to findings on the different level-one factors separately. To show and interpret the relative strength of the various findings, one would need to refer to studies with a true multivariate design, that is, those covering not only a multitude of countries but also a sufficient number of variables representing *each* of the first-level factors in the models. Because of the complexity of methodological and data requirements, none of the studies attempts to cover all of the first-level factors simultaneously, but a few studies in our literature review were able to cover a multitude of the factors mentioned above.

One of the few examples is OECD (2011), which makes an attempt to study the interactions between four groups of factors: (i) globalization (captured both by trade and financial openness); (ii) SBTC; (iii) institutional and regulatory reforms; and (iv) changes in employment patterns.[77] When explaining the relative weights of these factors within a common analytical framework,[78] the authors conclude that globalization (trade, FDI, financial liberalization) had little effect on wage inequality trends per se once institutional factors are accounted for. However, globalization processes put pressure on policies and institutional reforms to deregularize labor and product markets. Such institutional and regulatory reforms were primarily aimed at promoting growth and productivity, and while they had a positive effect on employment, at the same time they have been associated with increased wage inequality in many countries. What concerns the role of technology development in the period is that it was mostly beneficial for the highly skilled workers, a trend that resulted in larger wage disparities. However, increases in human capital (via mostly large-scale expansion of higher education in most OECD countries) offset much of the drive towards rising inequality.

Another example is Cornia (2012), who examined the explanatory factors of the declining inequality trends in Latin American countries. Among "proximate" causes of inequality, he investigated changes in both factorial and personal distributions of income caused by endowments of unskilled labor, human capital, physical capital, land and nonrenewable assets; their rates of returns also were taken into account. State intervention was measured by taxes and transfers received by households. Household-level income components enter the equation (similar to GIRE), together with macro-level variables such as dependency rates and activity rates. Overall inequality (measured by Gini coefficients) was decomposed into a weighted average of six factors (six different types of income). Results then were put into a broader framework, and changes in proximate causes are interpreted within the frame of changes in underlying causes (these include external conditions such as exports or capital flows, macrovariables related to the balance of payments, nonpolicy endogenous factors such as fertility and activity trends, dependency ratios, etc.), educational achievements and policy factors (related to taxes and transfers policies, wages, labor markets, economic and social policies, etc.). The major conclusion of the paper is that the decline in inequality in Latin America was most importantly due to the reversal of the skill premium (resulting from a massive increase of secondary enrolment), a decrease in the supply of unskilled labor, a return to

[77] In a second step, when moving from explaining individual earnings inequality to explaining household earnings and income inequality, the study adds two additional factors to the framework: (v) changes in family formation and household structures; and (vi) changes in tax and benefit systems. These have been identified as two of the key drivers of the increase in inequality up to the Great Recession, as the redistributive effectiveness tended to decline, mostly starting in the mid-1990s.

[78] For applying a joint framework for capturing the distribution effects on both wages and employment, the study uses a methodological approach proposed by Atkinson and Brandolini (2006b).

collective bargaining and an increase in minimum wages. Other factors such as the improvements in external economic conditions or the endogenous changes in dependency and activity rates played only a minor role in inequality reversal.

A third noticeable example for an attempt to create a broad based modelling of inequality change is Mahler (2010), who sought to explain the determinants and effects of government redistribution on inequality, mostly focusing on the role of taxes and transfers and on the distributive effect of wage bargaining institutions and minimum wages. He tested five alternative explanations from the literature: the median voter argument, the PRT, the political institutional approach, the labor unions approach and the globalization approach. Government redistribution was found to be positively related to pregovernment inequality (as the MR argument predicts), to the level of electoral turnout, to unionization rate and to the presence of proportional electoral systems. Further, a relatively egalitarian distribution of earnings was found to be positively associated with the degree of coordination of wage bargaining. On the other hand, no significant relationship has been found for the measures of globalization in his models.[79] The study also does not find support for the government partisanship hypothesis (share of cabinet positions held by Left parties).

These three examples are quoted here in more detail because they help show how far the various multivariate analyses can take us in understanding the relative weights of the various drivers of inequality. However, for a more encompassing GIRE-type specification and a proper test of it, still better data and larger country coverage are awaited.

19.6.2 Lessons on Methods and Models

We started this chapter with the aim to provide a thorough survey of what international (i.e., cross-country) studies can tell us about the drivers and underlying causes of income inequality with regard to levels and, in particular, trends. In the sections above, we were able to demonstrate how much progress has been made in terms of data availability and use for the countries in the joint set of the EU and the OECD (despite all remaining deficiencies of secondary data sets). A rich literature of studies of various drivers of inequality and their results have been discussed in the chapter. Yet, for the answers to some of the most important questions formulated at the outset, the jury is still out. These relate to

- the influence of the time coverage and geographical coverage of inequality data
- a more precise identification of the relative weights of factors (drivers) of inequality
- the comparability and accuracy of model estimates

[79] "Although the prospect that globalization will bid down social transfers and constrain earnings of low income groups looms large in the popular consciousness, it does not appear that a country's integration into the world economy seriously undermines government redistribution in the developed world" (Mahler, 2010, p. 529).

Below we discuss these three aspects in turn.

The articles reviewed in this chapter reveal that there have been quite spectacular developments in data infrastructures for the research on earnings and income inequality. Elements of this development can be summarized as follows:

- First and foremost, some new, large, comparative data collection exercises began. The most prominent one is the EU-SILC, produced annually for all of the member states of the EU and some non-EU countries. This data exercise encompasses a combination of ex ante and ex post harmonized data collection activities (Atkinson and Marlier, 2010).
- The collection of inequality variables in secondary data sets (most recently, the OECD Income Distribution Database, for example) has been accelerated and standardized and moved to annual reporting. In addition, some new secondary data sets have been built (of which the GINI project has most recently provided a rich data set for 30 countries and 30 years; Tóth, 2014).
- For some of the countries, a historical data collection exercise started, which contributes to a much better understanding of long-term trends in inequality (see, e.g., Atkinson and Morelli, 2014 or the long-run data series of the World Top Incomes Database developed by Alvaredo et al.)

In sum, the data situation improved greatly in the past few decades and even since the publication of Volume 1 of the *Handbook of Income Distribution* (Atkinson and Bourguignon, 2000). Simon Kuznets could now perhaps count on a situation where not 5% but maybe 50% of the analysis comes from data and only the other half (rather than 95% in 1955) of the analysis has to rely on speculation. Nevertheless, there are still deficiencies in the data front that impose serious limits on analysis and on a better understanding of the dynamics of inequality from a cross-country perspective.

While there are some data sets covering a large number of countries, there are a few truly longitudinal data sets covering long periods but only a few countries. However, researchers wishing to analyze inequality developments using comparable long-term series of country data will have to make serious compromises.[80] These types of compromises regard coverage (N), the number of data points (t) per country and their combinations as well.

The vast majority of studies reviewed is based on unbalanced panels because they cover different time periods for each country. That means that t has a variance across the cases. If this variance is nonrandom, the estimates may be biased. When missing years correlate in a systematic way with the dependent variable, estimates risk being biased. In addition, for income inequality estimates, annual time series are not available for most

[80] As an illustration, in the GINI project, involving hundreds of country experts and producing case studies for 30 countries (27 from the EU as of 2010) over 30 years between 1980 and 2010, only some two-thirds of all the possible cells of the 30×30 matrix could be filled with reasonably well-comparable Gini coefficients (Tóth, 2014).

countries and in general not in secondary data sets. Most of the studies summarized in Annex Table A19.1 look at a time period of about 20–30 years, but the number of observations per country differs greatly, from around 3 up to 20.

How serious the issue of unbalanced panels is also depends on the nature of the research question: for some tests of questions, a large N may compensate for a small *t*, for example, when testing the effect of institutional change (in which case the over-time variance in short periods will be negligible). In other cases, for example, when looking at the effect of macroeconomic changes (where year-to-year fluctuations may be not negligible), it may not.[81]

As we have shown in Sections 19.5.1–19.5.6 (roughly corresponding to the six major "diamonds" in Figure 19.1 representing six different groups of potential drivers of inequalities), studies of inequality identified significant effects of globalization and technical change, of political structures, of redistributive expenditures and some demographic composition changes. However, most models following the structure of Equation (19.3) (GIRE) are partial in the sense that they ask how variable group X affects inequality when controlled for variable groups Z or Q variables. This sometimes can misguide readers when interpreting the relevance of the results. All in all, in the literature there are rare attempts to provide weights to various significant factors; many leave complementary variable sets among the group of omitted variables or assume them to be absorbed by fixed effects.

As an example, studies analyzing the effects of globalization on inequality typically control for sectoral composition of the economy or sometimes for institutional variables (such as unionization or employment protection) but still leave out a great number of variables that could help control for demographic or education structure, for political processes or for redistribution. Similarly, analyses focusing on, for example, politics do account for party structures, electoral systems, voter turnout patterns and the like, sometimes controlling for demographic composition of societies, and so on. However, they also remain "rough," omitting too many variables (related to globalization, sectoral divisions, etc.) and thereby keeping a large part of the unexplained variance in the dark (or gray).

However, when trying to enrich the variable sets on the right-hand side of the GIRE, we run into problems similar to those of growth regressions. This does not come as a surprise because the structure of inequality regressions and those of growth regressions is similar, with just different left-hand variables. As indicated in the literature on economic growth regressions (see Mankiw, 1995; Temple, 2000; Eberhardt and Teal, 2009), part of the problem of inconclusiveness of results stems from a very simple fact: too small a number of countries, too many competing explanations and too short a time series with not many comparable definitions. Mankiw (1995) lists three of these

[81] For instance, the *U*-shaped inequality development in France between 1985 and 2010 (with the lowest point reached in 1998) requires more frequent year observations to perform meaningful econometric analysis.

problems: the problem of simultaneity, the problem of multicollinearity and the problem of degrees of freedom. For inequality regressions, each of these holds equally.

Simultaneity refers to the fact that right-hand variables are, in many cases, not exogenous but products of the same third (sometimes unobserved) factor, which determines inequality, and the chosen right-hand-side variable as well. This problem can also be called the endogeneity problem or reverse causality. Should we find that inefficient redistribution in a country fails to produce the expected inequality reduction, it might easily be that both government inefficiency and the large market income inequality are a product of a third factor, such as bad governance and or distrust in the given country (also on this issue see Robinson, 2009).

Multicollinearity has a similar origin. In many of the models the right-hand variables are correlated. A high level of taxes, for instance, will correlate with high levels of expenditures, especially in countries with higher levels of state employment (which in itself may have a lower level of inequality within this sector). Also, a higher share of more educated people may correlate with higher employment in education, where wage bargaining is more centralized. Inequality regressions need to face these multicollinearities, and researchers need to be innovative in trying to find proper ways to decrease the level of multicollinearity problems.

The third aspect is related to the potential number of explanatory variables. The trade-off here can be summarized as follows. For partial regressions, there may be too much unexplained variance left for the omitted variables. For more comprehensive regressions, the small number of observations limits the options. Given the fact that cross-country comparisons usually cover only a limited number of countries, the increase in the number of independent variables also is constrained. As Mankiw (1995) puts it, "there are too few degrees of freedom to answer all the questions being asked" (p. 306). For a better understanding of how inequalities evolve in a cross section of countries, more data points are needed—but for this we cannot have more countries, only time observations.

Furthermore, with the current amount of information at hand, not all of the complex mechanisms and channels that affect the distribution of earnings and incomes will show up in aggregate inequality regressions. Therefore, attempts to better specify the GIRE need to be complemented with more analysis of the constituent parts of these channels.

A final but important lesson relates to the disciplinary composition of inequality researchers. In our review we covered literature from economics, sociology and political science. Our most important lesson from this was that these disciplines have something to tell and to learn from each other. To share knowledge and discuss results, a common language is needed. As we have seen from scrolling though the literature, it is starting to exist.

As Atkinson and Brandolini (2009) put it, "valuable lessons can be learned but that we require: an integrated approach to theory and estimation; a proper specification of the data employed; and techniques to address the deficiencies of the underlying data" (p. 442). This will help decrease the level of speculation in inequality research—what Kuznets estimated to be 95% and we estimate now to be around 50% because of the fast development of inequality research in the past few decades.

Annex Table A19.1 Summary of multivariate analyses of determinants of cross-country differentials of within-country income distributions

Author, date	Geographical coverage, period and number of inequality observations	Data source for inequality measure	Dependent variable (inequality measure)	Explanatory variables and regressors	Estimation method	Findings with regard to causal factors of inequality	Other main findings
Globalization							
Spilimbergo et al. (1999)	34 Countries (21 OECD countries), 1965–1992, 320 observations	Deininger and Squire (1996 version)	Gini coefficient of personal income distribution	– Endowments: arable land per capita; capital per worker; skill intensity – GDP per capita in PPPs (and squared) – Endowment-corrected measure of trade openness	OLS	– Land- and capital-abundant countries: significant positive – Skill-abundant countries: significant negative – Trade openness (keeping constant factor endowments): significant positive – Trade openness (interacted with factor endowments): significant positive in skill-abundant countries but significant negative in capital-abundant countries	For the subsample of developing countries, coefficient of openness measure itself is negative but not significant
Heshmati (2004)	60 Countries (29 OECD countries); years between 1995 and 2000	WIID1	Gini coefficient of income (concept not specified) for most recent year and average across all available years	– Kearney composite index of globalization (covering economic integration, personal contact, technology and political engagement) – Regional dummies	Cross-sectional OLS	*Overall globalization index:* – Significant negative *Subcomponents of globalization:* – Personal contacts and technology (internet use): significant negative (especially personal contacts) – Economic integration: significant positive (but insignificant when average Ginis are used) – Political engagement: insignificant	Economic integration does not systematically lead to increased income inequality. Overall, globalization explains little of cross-country variations of inequality. Regional heterogeneity captures most of the variation

Mahler (2004)	14 OECD countries, 59 observations, 1980–2000	LIS	Gini coefficient of households, earnings	– Economic globalization: LDC trade (share in GDP of imports from LDCs), outbound investment flow as percentage of GDP, financial openness (14-point scale) – Domestic factors: ideological balance (5-point scale), electoral turnout, union density, wage coordination (5-point scale), unemployment rate, female participation rate	OLS, fixed effects	– Concerning earnings inequality: financial openness (+), electoral turnout (–), wage coordination (–). – Concerning fiscal redistribution: wage coordination (–) – Concerning disposable income inequality: union density (–), wage coordination (–)	–
Milanovic (2005)	129 Countries, 3 benchmark years (1988, 1993, 1998), 321 observations	WYD	Mean-normalised per capita household income deciles	– Trade openness (exports + imports)/GDP) – FDI/GDP *Controls*: – Financial depth (M2/GDP) – Democracy indicator *Extended version*: interest rate and government expenditure/GDP	Pooled cross-sectional OLS for each income decile; GMM instrumental variable estimation	– Increased trade openness: pro-rich in lower-income countries but pro-poor and middle incomes in higher-income countries (from around $8,000 PPP per capita) – FDI: no effect – Financial depth: increases low- and middle-income shares – Higher democracy index: increases middle-income shares – Higher government expenditures: pro-poor – Higher interest rates: pro-rich	Introducing regional dummy variables increases significance of results

Continued

Annex Table A19.1 Summary of multivariate analyses of determinants of cross-country differentials of within-country income distributions—cont'd

Author, date	Geographical coverage, period and number of inequality observations	Data source for inequality measure	Dependent variable (inequality measure)	Explanatory variables and regressors	Estimation method	Findings with regard to causal factors of inequality	Other main findings
Figini and Görg (2006)	107 Countries (22 OECD countries), 1980–2002, 664 (200) observations	UNIDO	Gini coefficient and Theil index of pay dispersion in the manufacturing sector	– Inward FDI stock/GDP, squared, and lagged *Controls*: – GDP per capita, secondary education enrolment, trade openness	– OLS fixed effects – GMM	*OECD countries*: – Inward FDI: significant negative – Education, trade: insignificant	– Nonlinearity of FDI effects is rejected when using lagged FDI and with GMM
IMF (2007)	51 Countries (19 OECD countries), 1981–2003, 271–288 observations	PovCal, supplemented by WIID2b (2007 version)	Gini coefficient and quintile shares of per capita income (not defined)	– Nonoil exports/GDP – Tariff liberalization (100-tariff rate) – Inward FDI-stock/GDP *Controls*: – ICT share in capital stock; credit to private sector/GDP; education (attainment and average years); agricultural and industry employment share	OLS, fixed effects	*Full sample*: – Technological progress: significant positive (largest contribution to increasing income inequality) – Globalization: small disequalizing effect, (equalizing) effect of trade offset by (disequalizing) effect of inward FDI – Financial depth (credit to private sector): significant positive *Subset of advanced countries*: – Technological progress: significant positive – Globalization: disequalizing (contributed somewhat more than technology); imports from DCs and inward debt significant negative but inward and especially outward FDI significant positive	Using quintile shares as dependent variables confirms findings, but estimates are less precise for tariff liberalization and technological progress

and Squire (2007)	1983–1999, ~170 observations (OWW) – ~90 countries, 1975–1999, ~170 observations (UTIP)	wages around the world (OWW) (ii) UTIP/UNIDO	of interoccupation wage inequality – Theil index of inequality for interindustry wages	tariff rate – Import-weighted indicator of presence of trade reforms in country's most important trading partners – GDP per capita – Union density – Coverage of collective agreements – Social expenditures/ GDP	and regressors in first differences	*inequality*: – Decrease in tariff rate: significant positive in poorer countries, but significant negative in richer countries – Labor market conditions: insignificant *Interindustrial wage inequality*: – Decrease in tariffs rate significant positive in poorer countries, particularly in countries with a high density of trade unions; but negative in richer countries – Social expenditures: insignificant	on both interoccupational and interindustry inequality reverses at around the world median level of income ($4000 in 1995 PPPs)
Bertola (2008)	51 countries (14 OECD countries); 1970–2000; 467 observations	WIID1 (2007 version)	Gini coefficient of household income (generally net income; observations for gross income are controlled by dummy variables)	– Trade openness (exports + imports)/ GDP – Share of government in GDP – Ratio of private credit to GDP – With and without controls for GDP per capita	OLS (fixed effects)	– Trade openness: significant positive – Financial development: significant positive – Government spending: significant negative (estimated by fixed effect regressions)	Government spending is less effective (i.e., less significant negative effect on inequality) if openness is high
Dreher and Gaston (2008)	– 100 countries (27 OECD countries) and 411 (129) observations for earnings – 100 countries (26 OECD countries) and 340 (110) observations for income – 1970–2000, averages of 5 years	UTIP/UNIDO	– Theil index of individual earnings – Gini coefficient of gross household income (estimated from earnings inequality above)	– Synthetic globalization index (KOF) and separately for economic, political and social globalization – Lagged dependent inequality variable *Controls*: – GDP per capita and its square – Democracy index	OLS fixed effects and GMM	– Globalization (overall index): significantly positive on earnings and income inequality in OECD countries (GMM: only income inequality) – No systematic evidence for three subdimensions of globalization – Lagged dependent: highly significant positive	– Greater democracy has not decreased earnings or income inequality – Results robust when including additional variables: demography (dependency ration, population growth) and size of government (government consumption expenditure) – Replacing time dummies with a time trend in all models leads to more poorly fitting models

Continued

Annex Table A19.1 Summary of multivariate analyses of determinants of cross-country differentials of within-country income distributions—cont'd

Author, date	Geographical coverage, period and number of inequality observations	Data source for inequality measure	Dependent variable (inequality measure)	Explanatory variables and regressors	Estimation method	Findings with regard to causal factors of inequality	Other main findings
Gourdon et al. (2008)	Dataset 1: 61 countries (22 OECD countries); 1980–2000; 198 observations Dataset 2: 55 countries (20 OECD countries); 1988–1998; 146 observations	Dataset 1: WIID1 (2005 version) Dataset 2: WYD	Gini coefficient of income (generally net household income; observations for gross income, personal income and expenditure are controlled by dummy variables)	– Trade openness: lagged ratio of tariff revenues to imports – Relative factor endowments *Controls*: – Log GDP per capita in PPPs (if no interaction with factor endowment) – Other controls (inflation, education, ethnicity)	OLS (fixed effects)	– Trade openness (changes in tariffs): significant positive *When interacted with factor endowments*: – Trade openness: significant positive in capital-abundant and highly skill-abundant countries – Macro stability (reduction in inflation): significant negative	– Analyzing more detailed data (decile data from WYD) but over a shorter time periods shows similar results but estimates often lack precision. – All studies should control for the source of inequality data via dummies and for omitted variable bias via FE.
ILO (2008)	16 OECD countries, 1978–2002, 175 observations	WIID2b, supplemented by LIS, PovCal, SEDLAC and TransMONEE	Gini coefficient of income (not specified but probably both net and gross)	– Globalization: inward FDI, tariff liberalisation, capital account openness – Institutions: union density, collective bargaining coordination – Other controls: ICT share, education years, credit to private sector, public social expenditures	OLS or FGLS (not specified), fixed effects	– Institutions: insignificant – Social expenditures: significant negative – Technology (ICT share): significant positive – Capital openness: significant positive – Education: significant negative	– FDI significant positive only when technology variable is not taken into account – PCA analysis suggests that from 1990s institutions forfeited capacity to reduce market inequality directly and retained indirect influence by virtue of the size of welfare state
Çelik and Basdas (2010)	Five developed countries 1995–2007 (FR, GE, NL, UK, US); 5 developing countries 1995–2006 (AG, BZ, CZ, HU, PL); 6 miracle countries 1990/1995–2005 (CH, IN, KO, MA, SI, TH) → 8 OECD countries	UNU-WIDER (WIID2), US Census Bureau (USA), EUROSTAT and other statistical departments	Gini index (household, income)	FDI inflows, FDI outflows and trade openness	OLS	– Developed countries: FDI inflows, outflows significant negative, trade openness positive – Developing countries: FDI inflows negative, trade openness positive – Miracle countries: FDI inflows positive, trade openness negative	Results cannot be generalized.

OECD (2011), Part I	22 OECD countries, 1979–2008, 333 observations	– OECD earnings database – LIS (for overall earnings distribution among working-age population)	– Interdecile ratios of full-time earnings: D9/D1, D9/D5, D5/D1 – Employment rate	– Trade exposure and subcomponents – FDI restrictiveness, inward and outward FDI, FPI, cross-border assets, private credits – Business sector expenditure on R&D – Institutional variables: union coverage, EPL (regular and temporary), tax wedge, UI replacement rate of low-wage workers, minimum wage, PMR *Controls*: – Education, sectorial employment share, female employment rate, output gap	OLS fixed effects	*Earnings inequality among workers*: – Globalization (trade and financial integration): insignificant – Technology: significant positive – Institutions/ regulations: significant negative – Education, female employment: significant negative *Overall earnings inequality among working-age population*: – Globalization (trade and financial integration): insignificant – Technology: positive – Education: negative – EPL (temporary): negative – Other institutions/ regulations: undetermined, sign depends on assumption of reservation wage	*Earnings inequality among workers, effects of subaggregates*: – Imports from low-income DCs: positive in low-EPL countries and negative in high-EPL countries – Inward FDI: negative, outward FDI: positive – Distributional effect of EPL is entirely driven by EPL for temporary contracts – Minimum wage (smaller sample): significant negative *D9/D5 and D5/D1 effects*: – FDI deregulation: reduces dispersion at bottom half and widens it at top – Technology: positive for D9/ D5 only – PMR and EPL: positive for D5/ D1 only – Union coverage: negative for D9/ D5 only – Education and female employment: negative for both D5/D1 and D9/D5

Continued

Annex Table A19.1 Summary of multivariate analyses of determinants of cross-country differentials of within-country income distributions—cont'd

Author, date	Geographical coverage, period and number of inequality observations	Data source for inequality measure	Dependent variable (inequality measure)	Explanatory variables and regressors	Estimation method	Findings with regard to causal factors of inequality	Other main findings
Zhou et al. (2011)	62 countries (24 OECD countries), benchmark year 2000	WIID2b (2004 version)	Gini coefficient of net income (observations on expenditures were increased by 5 points, on gross income decreased by 7.5 points)	– Globalization: equally weighted Kearney index and principal component index – Education level (HDR education index) – Urbanization level	Cross-sectional OLS	– Both overall globalization indices: significant negative – Education: significant negative – Results of globalization are robust to inclusion of education and urbanization	*Subcomponents of globalization*: – International travel and Internet user: significant negative – Trade: significant positive – FDI: insignificant
Cassette et al. (2012)	10 OECD countries (AS, DK, FI, FR, GE, JP, NE, SW, UK, US), balanced panel; 1980–2005; 220–240 observations	OECD earnings database	Interdecile ratios of individual earnings: D9/D1, D9/D5, D5/D1	– Trade openness: total, goods and "other" services *Controls*: – FDI stock – Education (average years of schooling) – GDP per capita – Inflation – Technology (ICT capital/total capital stock) – Institutions (union density and concentration, bargaining level)	Error correction model regression	*Long-run effects*: – Trade in goods: significant positive on D9/D1 and D5/D1 – Trade in services: significant positive on D9/D1, D9/D5 and D5/D1 – FDI, GDP/capita: significant positive *Short-run effects*: – Trade in goods: significant positive on D9/D1 and D5/D1 – Trade in services: insignificant – FDI, GDP: insignificant	Education has a negative effect on inequality (but coefficient not always significant) Union density and union concentration: significant and negative
Faustino and Vali (2012)	24 OECD countries; 1997–2007; 230 observations	WIID2 (2008 version) (missing values imputed)	Gini coefficient of income	– Trade openness: (exports + imports)/GDP – FDI (net inflows/GDP) *Controls*: – GDP/capita, unemployment, LTU, inflation, number of companies	– OLS fixed effects – GMM	*OLS*: – Inward FDI significant positive – Trade openness significant negative *GMM*: – Inward FDI insignificant – Trade openness significant negative	*OLS*: – GDP/capita, unemployment and inflation significant positive, other controls insignificant *GMM*: – GDP/capita significant positive, other controls insignificant

Institutions							
De Gregorio and Lee (2002)	22 countries (1965), 49 countries (1990) (18 OECD)	IMF Government Finance Statistics Yearbook	Gini coefficient and quintile shares (household, income)	Educational inequality, educational attainment, log of GDP/capita, square of log of GDP per capita, social expenditure/GDP, regional dummies	OLS	Nonlinear relationship between educational attainment and educational inequality (inverted *U*-shape)	–
Beck et al. (2004)	For changes in the distribution of income: 52 developing and developed economies with data averaged over the period 1960 to 1999; For changes in poverty: 58 developing countries with data over the period 1980 to 2000	World Development Indicators, Dollar and Kray (2002), PovCal Net	Changes in 4 separate dependent variables: (i) changes in poverty (change in income of each economy's poorest 20%); (ii) changes in income distribution (Gini coefficient); (iii) growth rate of the percentage of population living under 1$ a day (and 2$ in robustness tests); (iv) growth rate of the Poverty Gap (=weighing by distance from the 1$ level)	GDP % of private credit by financial intermediaries to private firms + GDP growth; Instrumental variables: legal origin of the country, latitude of the capital city, natural resource endowments; plus for inequality models: initial (1960) avg schooling, inflation, trade openness; plus for poverty models: initial poverty level	OLS, 2SLS	**1.** Financial development alleviates poverty and reduces income inequality **2.** Countries with better-developed financial intermediaries experience faster declines both in poverty and income inequality	–
Checchi and Garcia-Penalosa (2005)	16 OECD countries, 1960–1996, 210 observations	Deininger and Squire (1998 version) Brandolini (2003)	Gini coefficient of personal incomes	– Labor share – Wage dispersion – Unemployment rate – Unemployment benefit	OLS and IV, fixed effects SLS regressions	– Labor share: significant negative – P9/P1 ratio: significant positive – Unemployment rate: significant positive (insignificant in OLS) – UB benefit: significant negative (indirectly through labor share in SLS) *Reduced form equation*: capital/labor ratio and education have strongest correlation with inequality, followed by union density, tax wedge and UB; minimum wage marginally significant	Labor market institutions (union density, minimum wage, unemployment benefit) are essential determinants of labor market outcomes: labor share, wage differentials, unemployment rates

Continued

Annex Table A19.1 Summary of multivariate analyses of determinants of cross-country differentials of within-country income distributions—cont'd

Author, date	Geographical coverage, period and number of inequality observations	Data source for inequality measure	Dependent variable (inequality measure)	Explanatory variables and regressors	Estimation method	Findings with regard to causal factors of inequality	Other main findings
Weeks (2005)	7 OECD countries (AS, CN, GE, JP, SW, UK, US); 1980–1998; 61 observations	WIID1	Gini coefficient of gross personal income	– Current public expenditure share in GDP – Unemployment rate – Union density rate	OLS (fixed country effects)	– Union density: significant negative – Public expenditure: significant negative – Unemployment: significant positive	Applying the model to two countries with annual time series (UK, US) yield the same strong significance for union density but unemployment and government expenditure (UK only) become insignificant
Carter (2007)	39 countries (20 OECD), 104 observations at all levels of economic development	WIID2b	Gini coefficient	– Economic freedom – Per capita income – Political rights – Civil liberties *Controls*: – Years of education; percentages of population under 15; over 64; urban; employed in industry; employed in services – Quadratic specification also included	OLS with robust standard errors	Economic freedom lowers equality by reducing income distribution towards the poor – However, if controls and fixed effects are omitted, the estimated trade-off between inequality and economic freedom disappears	–
Checchi and Garcia-Penalosa (2008)	16 OECD countries, 1969–2004, 82 observations	LIS	Gini coefficient of 3 equivalized income definitions for the working-age population: factor income, gross income, disposable income	– Institutions: union density, unemployment benefit, EPL, wage coordination, minimum wage, tax wedge *Controls:* – Demography: age of head of household, age of spouse – Tertiary education – Other controls: female employment, investment, openness	OLS, fixed effects	*Factor income inequality*: – Institutions insignificant, except tax wedge (significant positive) *Gross and disposable income inequality*: – Unemployment benefit, EPL: significant negative – Tax wedge: significant positive	– Trade-off of unemployment benefit and EPL: both lower inequality but increase unemployment (EPL only without fixed effects) – Weaker effect of institutions on factor income than on disposable income inequality

Beramendi and Cusack (2009)	13 countries (all OECD), 41 observations, 1978–2002 (LIS 5-year time periods)	LIS	Gini coefficient for market income inequality, wage inequality and disposable income inequality	– First model (wage inequality): number of manufacturing workers, imports from the Third World (percentage of GDP), female labor force participation rate, proportion of at least college education, union density, government partisanship, economic coordination, interaction of the last two – Second model (market income inequality): wage inequality, stock market capitalization, percentage of population in retirement age – Third model (disposable income inequality): union density, economic coordination, government partisanship	OLS (robust standard errors and panel-estimated standard errors)	First model (effects on wage inequality): female participation (+), percentage of college education (+), union density (−), economic coordination (−), interaction of economic coordination and partisanship (−) Second model (market-based income inequality): stock market capitalization (+), pension-age population (+) Third model (disposable income inequality): market income inequality (+), union density (−), economic coordination (−) Left government inheritance (−)	–
Carnoy (2011)	20 countries (3 OCED); 1960–2003	WDI	Gini coefficient of household, income, highest 20%, lowest 20%	Trends of inequality, distribution of education, private and social returns to education, ratio of public spending	Trend analysis (no regression)	Higher education: greater inequality	Logical chain: higher education + → differentiation + → better (and richer) students to better universities → returns of education differentiate → greater inequality
Golden and Wallerstein (2011)	16 OECD countries (AS, AT, BE, CN, DK, FI, FR, GE, IT, JP, NE, NO, SW, CH, UK, US);	OECD earnings database	Interdecile ratio of individual earnings: D9/D1	*First differences over 5-year periods*: – Deindustrialization: share of industrial employment	Weighted OLS; separate regression models for 1980s and 1990s; IV (independent variables); extreme	1980s: – Union density and centralisation: negative and highly significant	Determinants of earnings inequality are different in 1980s (institutions) and 1990s (trade with LDCs and social

Continued

Annex Table A19.1 Summary of multivariate analyses of determinants of cross-country differentials of within-country income distributions—cont'd

Author, date	Geographical coverage, period and number of inequality observations	Data source for inequality measure	Dependent variable (inequality measure)	Explanatory variables and regressors	Estimation method	Findings with regard to causal factors of inequality	Other main findings
	1980–2000; around 220 observations			– Globalization: total trade; trade with LDCs – Institutions: union density; centralization *Controls*: – Migrants share in population; Right parties share in parliament; social insurance expenditures/GDP; unemployment rate; female labor force participation	bounds analysis to test robustness	– Trade, deindustrialization: positive but insignificant – Other controls: insignificant 1990s: – Trade with LDCs: positive and significant – Social insurance expenditures: negative and significant – All other regressors and controls: insignificant	expenditures), but in neither period is deindustrialisation significant
Muinelo-Gallo and Roca-Sagalés (2011)	Unbalanced panel of 43 upper-middle- and high-income countries for 1972–2006	WIID2b (Gini coefficients)	(Log of) Gini coefficient (5-year averages) of income	– Civil liberties – Education inequality – Growth – Public investment – Current public expense – Direct taxes – Indirect taxes – Disposable income (dummy) *Control*: – Dummy for various data sources	– OLS (pooled, one-way random effects models with temporal dummies)	– Increase in civil liberties index reduce income inequality – Increase in educational inequality increase inequality – Current public expenditure has significant and sizeable negative effect on inequality – The direct effect of public investment is not significant (though indirect effects are shown) – Direct taxes have negative and significant (though small) effect – economic growth has a significant negative effect on inequality	– Data source dummy is significant on Gini – Public current expenditures and direct taxation robust in sensitivity estimates

Rueda and Jonas Pontusson (2000)	16 OECD countries (1973–1995)	OECD earnings database	Interdecile ratio of individual earnings: D9/D1	Union density, centralization of wage bargaining, the public sector's share of employment, partisan composition of government, social vs. liberal market economy, social spending/GDP, collective bargaining coverage and employment protection	OLS fixed effects	Union density, centralization of wage bargaining, the public sector's share of employment, and government partisanship: significant equalizing effect	The effect of institutional variables differs between the various variations of capitalism (liberal market economies [LMEs] and social market economies [SMEs]).
Bradley et al. (2003)	61 observations from 19 OECD countries, 1967–1997	LIS	Pre-tax, pre-transfer income distribution and proportional reduction in inequality from pre- to post-taxes and transfer inequality (based on household, income Gini coefficients)	Welfare generosity; Leftist party and Christian democratic party share of government; veto points; union membership; bargaining centralization/ corporatism +globalization (4 measures); economic development (GDP per capita and agricultural employment); deindustrialization (industrial employment); secondary school enrolment, vocational education; percentage of the unemployed, of female labor force participation, of female-headed households and of population under the age of 15	OLS	Unemployment (+), female-headed family (+) and union density (−); secondary education not significant	Effects on governmental redistribution: taxes, transfers, Leftist government The hypothesis that the magnitude of taxes and transfers has a strong effect on the reduction of inequality is confirmed.
Kenworthy and Pontusson (2005)	11 countries (1979–2000) (all OECD)	LIS	Gini coefficient (household income)	–	Trend analysis	–	Redistribution increased, but the effects were hidden by the even more increasing market inequality. Voter turnout can explain the redistribution paradox (the more inequality the less redistribution).

Continued

Annex Table A19.1 Summary of multivariate analyses of determinants of cross country differentials of within country income distributions—cont'd

Author, date	Geographical coverage, period and number of inequality observations	Data source for inequality measure	Dependent variable (inequality measure)	Explanatory variables and regressors	Estimation method	Findings with regard to causal factors of inequality	Other main findings
Iversen and Soskice (2006)	14 countries, 61 observations, 1967–1997 (all OECD)	LIS	Gini coefficient (household, income) before and after tax	Government partisanship, electoral system, pre-tax and tax inequality, constitutional veto points, unionization, voter turnout, unemployment, real per capita income, female labor force participation	OLS	*Effects on redistribution*: Right government (−), veto points (−), unionization (+), female labor force (−), GDP (−), unemployment (+), voter turnout (+)	–
Rueda (2008)	16 OECD countries (1973–1995)	OECD Economic Outlook	Gini coefficient (earnings, individuals)	Cabinet partisanship, unemployment, LDC trade, female labor force participation, private service employment + international and financial openness, government debt, unemployment, GDP growth	OLS fixed effects	*In case of low corporatism*: Left governments increase government employment and minimum wages, but reduce welfare state generosity, with following effects on inequality: employment (significant negative); minimum wages (significant negative); generosity (positive, not significant) *In case of high corporatism*: Left governments reduce government employment, minimum wages and welfare state generosity, which have the following effect on inequality, respectively: − (significant), + (not significant), − (not significant). *In general*: Left governments do not have a significant effect on inequality.	–
Iversen and Soskice (2009)	16 OECD countries, 1880–1990	LIS, Cuzack (2003), Cusack and Fuchs (2002)	Gini coefficient (individual, earnings) before and after tax	Electoral system, degree of nonmarket economic coordination + size of the electorate, size of the elderly population, GDP per capita; interactions of electoral system with decade dummies	OLS (fixed effects)	Electoral system and partisanship: significant effects on redistribution	Proportional representative electoral systems: positive effect on social spending shocks

Mahler (2010)	59 observations in 13 countries (all OECD), 1979–2000	LIS	Gini coefficient and percentiles (upper/lower, upper/middle and middle/lower percentile ratios)	Dependent variable: government redistribution Regressors: electoral turnout, skewness of turnout by income, distribution of pre-government income + share of the population over 65, ideological balance of the governing cabinet, share of imports from less developed countries in GDP, share of outbound FDI in GDP, and a measure of the openness of a country's economy to global financial flows, electoral disproportionality, competitiveness of elections, voter registration	OLS	Voter turnout is positively correlated to government redistribution even after controlling for pregovernment inequality	–
Pontusson and Rueda (2010)	10 OECD countries, 1966–2002	Atkinson (2007)	Share of total income accounted for by the top 1% of income earners (individual)	Voter turnout, effective number of parties, dummy for the existence of Left-wing competitors, median voter, union density	OLS	Left party position reduces inequality, if median voter mobility is high	–
Redistribution							
Fuest et al. (2009)	26 EU countries (except Malta), 2007	EU-SILC	Squared coefficient of variation, GE(2)	–	Measurement of the contributions of social policy instruments to redistribution with two methods (standard and decomposition), cluster analysis	– Findings from the standard approach: benefits are the most important source of redistribution – Findings from decomposition approach: taxes are the most important ones (differences can be understood through focusing on the main goal of both approaches)	According to cluster analysis, new EU member countries do not form a distinct group, central eastern European countries tend to show similarities with continental welfare states, whereas the Baltic countries with the Mediterranean ones

Continued

Annex Table A19.1 Summary of multivariate analyses of determinants of cross-country differentials of within-country income distributions—cont'd

Author, date	Geographical coverage, period and number of inequality observations	Data source for inequality measure	Dependent variable (inequality measure)	Explanatory variables and regressors	Estimation method	Findings with regard to causal factors of inequality	Other main findings
Mohl and Pamp (2009)	23 OECD countries, 1971–2005	LIS	5-Year averages of cumulative share-gains of the first, the first to second and the first to fifth deciles and share-gain of second to eighth deciles	– Overall government expenditures – Government social expenditures – Social transfers ratio (average transfers per total disposable income) – Unemployment expenditures – Health expenditures – Gini – Percentiles ratios (P90/P50, P50/P10, median-to-mean ratio) – Left government – Disproportionality of the electoral system – Voter turnout *Controls*: – GDP growth – Unemployment rate – Population 65+	Panel regressions ($t=7$, $N=23$) with various robustness checks, two-step system GMM	– At very high levels of inequality the positive relationship between inequality and redistribution is reversed (nonlinear relationship) – Redistribution is driven by the P90/P50 ratio and targeted at the middle class (Director's law)	–
Afonso et al. (2010)	26 OECD countries; year around 2000 and average for period 1995–2000	WIID, supplemented by OECD and LIS	– Gini coefficient of household disposable income – Income share of bottom 40% – Per capita income of bottom 20% in PPPs	– Redistributive social spending (transfers, subsidies) – PIT – Education achievement (PISA) – Education spending – Unemployment – GDP per capita in PPPs	– Cross-sectional OLS – DEA for assessing efficiency of public spending – Tobit regressions to capture exogenous nondiscretionary factors in explaining spending efficiency	– Redistributive social spending: highly significant equalizing the distribution (all three inequality indicators) – Education achievement (in particular maths): significantly equalizing – Education spending and PIT: not significant – Only high social spending coupled with good education reduces inequality (Gini)	– DEA suggests low efficiency of public spending with regard to inequality in some southern and continental European and high efficiency in some Nordic countries – Tobit analysis suggests strong indirect role of institutions on distribution, being significantly correlated with spending efficiency

						– Higher social spending, higher per capita GDP and lower unemployment: associated with higher income of bottom quintile	
Goudswaard and Caminada (2010)	25 OECD countries (as in OECD 2008)	OECD (2008) data on income inequality	Redistribution (reduction of Gini coefficient of market income to Gini of disposable household income)	– Public social transfers (and elements such as pensions, active labor market benefits, unemployment benefits) – Spending on public services (health) – Private (pension) social expenditures	OLS	– Total public expenditure on redistribution: positive – The effect of total social expenditure (public and private) on redistribution: weaker but positive – At program level, pensions have larger effect than unemployment and labor market programs – Excluding health expenditure does not significantly affect the above – Private pensions expenditure on redistribution: negative	
Jesuit and Mahler (2010)	12 OECD countries, 52 observations, 1979–2004	LIS	Gini coefficient of household income	–	Estimating net redistribution (difference between pregovernment and postgovernment Ginis) with three different methods: standard, only pre-pension society, life-cycle based	Intra-individual redistribution arises mostly from government taxes and transfers, but its rate is quite different depending on pension systems and other social policies	Using the alternative method described in the article results in lower values for redistribution than based on the conventional measure
Niehues (2010)	24 EU countries (21 OECD), 183 observations, 1993–2006	European Community Household Panel (1993–2000), EUROMOD (2001), EU-SILC (after 2003)	Gini coefficient of equivalised disposable income	Social spending + lagged dependent variable, macroeconomic factors (GDP per capita, and GDP per capita squared), socioeconomic factors (dependency ratio, percentage of the	Nonparametric analysis, stochastic kernel on dimensional graph + GMM	– Social spending: negative (especially social benefits, unemployment benefits and old age survivor benefits, and not health benefits), – GDP (inverse U-shaped),	–

Continued

Annex Table A19.1 Summary of multivariate analyses of determinants of cross-country differentials of within-country income distributions—cont'd

Author, date	Geographical coverage, period and number of inequality observations	Data source for inequality measure	Dependent variable (inequality measure)	Explanatory variables and regressors	Estimation method	Findings with regard to causal factors of inequality	Other main findings
				population aged 25–64 with at least secondary education), union density, dummy for post-communist countries		post-socialism: negative Second-order effects are found to be nonsignificant	
Immervoll and Richardson (2011)	14 OECD countries, mid-1980s to mid-2000s	LIS	Gini (disposable, market income)	Redistribution as instrument to reduce inequality	Trend analysis, decomposition	–	– Tax-benefit systems are less effective than they were in the 1980s (despite the fact that they became more redistributive over the whole period). – In general, benefits decreased in real terms, although they were still the major drivers of redistribution. – Taxes contributed less to redistribution. – Redistribution strategies based on government transfers have to be complemented by employment policy
Doerrenberg and Peichl (2012)	Panel of OECD countries (unspecified)	– LIS – WIID – University of Texas Inequality Project (UTIP) – Penn U tables – World Tax Indicators (Sabirianove-Peter et al., 2010) – OECD statistics – WB WDI database	– Gini coefficient of household incomes – Regression estimates of household income inequality from wage inequality data	*Dependent variables (lagged 1 year)*: – Government spending (Penn) – Public total social expenditure (OECD Soc exp) – Degree of tax progressivity (WTA) *Controls (lagged 1 year)*: – GDP per capita – Squared GDP per capita – Trade openness – Inflation rate	– OLS with country and year fixed effects – 2SLS for instrumental variables	– 1% Increase in government spending decreases inequality by 0,3% – Tax progressivity is insignificant – Social expenditures are more efficient in inequality reduction than taxes – There are indications of second-order effects	– No significance of GDP (as a control variable) for the results – Inflation: slight negative effect – Union density: strong effect on inequality

				– Unemployment rate – Union density – Higher education levels – Index of globalization *Instrumental variable*: – 1981 level of policy variables (government spending, government social experiments, progressivity) extrapolated			
Structural and Macro							
Alderson and Nielsen (2002)	16 OECD countries, 1967–1992, 192 observations	Deininger and Squire (1996 version)	Gini coefficient of gross income	– GDP per capita (and squared) – Sector dualism (shift of employment out of agriculture) – Sector bias (share of labor force in agriculture) – Natural rate of population increase – Secondary school enrolment – 3 Globalization variables: DI outflow/labor force, southern import penetration/GDP; net migration rate – 3 Institutional variables: union density, wage setting coordination, decommodification – Female labor force participation – Period indicators (1970s and 1980s)	Random effects	– Sector bias: significant positive – All 3 globalization indicators: significant positive – All 3 institutional indicators: significant negative – Female labor force participation: significant positive – Secondary school enrolment: significant negative – Other controls: insignificant	– Only modest evidence for inequality trend being inherently linked to postindustrial development – When calculating relative contributions of factors, sector bias has the strongest effect (positive), followed by union density and decommodification (negative), southern import penetration and DI outflow (positive) – When estimating the maximum longitudinal effect (within single countries), sector bias is still dominant, but followed by southern import penetration and FI outflow

Continued

Annex Table A19.1 Summary of multivariate analyses of determinants of cross-country differentials of within-country income distributions—cont'd

Author, date	Geographical coverage, period and number of inequality observations	Data source for inequality measure	Dependent variable (inequality measure)	Explanatory variables and regressors	Estimation method	Findings with regard to causal factors of inequality	Other main findings
Rohrbach (2009)	19 OECD countries, 1970–1999, 225 observations	UTIP/UNIDO (EHII data set)	Gini coefficient of gross income (estimated from pay data)	– Sector dualism (knowledge sector wage differential) – Sector bias (employment share in knowledge sector) *Controls*: – Average years of schooling – Natural rate of population growth – Union density – Trade openness (trade/GDP constant prices)	Error Correction Model regression	– Sector bias (income differential in knowledge sector): significant positive – Sector dualism (differential between knowledge sector and other sectors): insignificant	– Average years of schooling lower inequality, in all specifications – No significant effects of demographic change and trade openness – Union density increases inequality
Social Structure							
Brady (2006)	18 Developed countries (mostly OECD)	LIS	Poverty head count Poverty intensity	– Manufacturing employment – Agricultural employment – Female labor force participation, the Elderly population – Children in single-mother families *Control variables*: – Economic development, welfare state	Random effects models and counterfactual simulations	– Manufacturing employment and female labor force participation significantly reduces poverty headcount – Share of elderly population and share of children in single mother families increases poverty headcount – Agricultural employment has no effect on headcount. – None of the explanatory variables are significant for poverty intensity	Welfare state has, in general, larger effect on poverty reduction than any of the structural explanatory variables. Economic development is mostly insignificant for (relative) poverty.
OECD (2011), chapter 5	23 OECD countries, mid-1980s to end of 2000s	LIS	Gini coefficient and D9/D1 ratio (of disposable income in 12 countries and of gross incomes in 11 countries)	– Dispersion of male earnings, male employment rates – Female employment rates – Assortative mating (earnings correlation of spouses)	Conditional reweighting and decomposition (see Chen et al., 2013b for methods)	– Main contributor to household earnings inequality: men's earnings dispersion – Increase in female employment has an equalizing effect in all countries – Change in men's employment has little	– Assortative mating has increased in OECD countries – There is a considerable heterogeneity in the size of unexplained increase in inequality

				– Household composition (five household types)		effect on household earnings inequality in all but three countries – Assortative mating and household structure changes has inequality increasing effects, but are less sizeable than employment effects	
General							
Li et al. (1998)	49 developed and developing countries between 1947 and 1994, 573 observations	Deininger and Squire (1996) data set	Gini coefficients (for inequality determinants models: averaged over 5-year periods)	*Dependents for testing cross-country variance*: – Country and years to measure cross-country and over-time variance *Dependents for determinants of inequality*: – Political economy variables (political freedom and initial secondary schooling) – Credit market imperfections (measured by land distribution and financial market development index) *Controls*: – Various definitions of Gini (income/consumption, etc.)	– Analysis of variance – OLS with dummies – OLS with instrumental variables	– 90% of total variance of Ginis is explained by variations across countries as opposed to intertemporal variation – 7 of 49 countries show significant decline, 10 show (small) increase in Gini – 65% of the sample of countries show no clear time trend – Financial market imperfection has larger effect on inequality than the political economy variables – Both effects are stronger for the lower 80% than for the top 20% – A more egalitarian distribution of land decreases inequality (benefits the poor more) – Expansion of political liberty, of secondary education and of financial market improvements benefits all and contributes to inequality decrease	–

Continued

Annex Table A19.1 Summary of multivariate analyses of determinants of cross-country differentials of within-country income distributions—cont'd

Author, date	Geographical coverage, period and number of inequality observations	Data source for inequality measure	Dependent variable (inequality measure)	Explanatory variables and regressors	Estimation method	Findings with regard to causal factors of inequality	Other main findings
Cornia (2012)	14 Countries in Latin America, 1990–2009 (2 OECD)	IDLA database	Gini coefficient (household, income)	– External conditions (international terms of trade, migrant remittances, and FDI); – The rate of growth of GDP per capita; – Changes in exogenous factors (the dependency rate and the activity rate); – The distribution of human capital among workers (the ratio of changes over time in the number of adults with secondary and tertiary education divided by changes over time in the number of those with primary or no education); – Fiscal policies (the ratio of direct to indirect taxes, and public expenditure on social security/GDP); – Labor market policies (the minimum wage interacted with the share of formal sector workers); – Macroeconomic policy (the real effective exchange rate and its square); (viii) political variables (the dummies "social democratic" and "radical-populist" and the Polity2 index, which measures the quality of democracy)	OLS, 3SLS, GMM	*Impact on Gini*: Terms of trade (−), remittances (not significant), FDI (+), GDP/c growth rate (−/not significant), dependency ratio (not significant), labor force participation (not significant), education (−), taxes (−), public expenditure (−), exchange rate (−), exchange rate squared (+), minimum wage (−), political variables (−), lagged Gini (++)	–

ACKNOWLEDGMENTS

The authors are grateful to Tony Atkinson, François Bourguignon and Andrea Brandolini for their useful comments and suggestions on earlier drafts of this chapter. The authors also thank Wen-Hao Chen, Tim Goedemé, Alexander Hijzen, Márton Medgyesi and Pieter Vanhuysse for comments and advice on the literature. Numerous comments at the Paris author's conference organized in preparation of this book in April 2013 are much appreciated. The authors also thank Anna B. Kis and Eszter Rékasi for their research assistance. They have no responsibility for any remaining errors. The views expressed are not necessarily those of the institutions with which the authors are affiliated.

REFERENCES

Acemoglu, D., 2002. Technical change, inequality, and the labor market. J. Econ. Lit. XL, 7–72.
Acemoglu, D., 2003. Labor- and capital-augmenting technical change. J. Eur. Econ. Assoc. 1 (1), 1–37, 03.
Acemoglu, 1988. Why do new technologies complement skills? Directed technical change and wage inequality. Q. J. Econ. 113, 1055–1090.
Adelman, I., Morris, C.T., 1973. Economic Growth and Social Equity in Developing Countries. Stanford University Press, Stanford, CA.
Afonso, A., Schuknecht, L., Tanzi, V., 2010. Income distribution determinants and public spending efficiency. J. Econ. Inequal. 8 (3), 367–389.
Aghion, P., Howitt, P., 1998. Endogenous Growth Theory. MIT Press, Cambridge, MA.
Aghion, P., Caroli, E., Garcia-Penalosa, C., 1999. Inequality and economic growth: the perspective of the new growth theories. J. Econ. Lit. 37 (4), 1615–1660.
Alderson, A.S., Doran, K., 2013. How has income inequality grown? The income distribution in LIS countries. In: Gornick, J., Jantti, M. (Eds.), 2013: Economic Disparities and the Middle Class in Affluent Countries. Stanford University Press, Stanford, CA, pp. 51–74.
Alderson, A., Nielsen, F., 2002. Globalization and the great u-turn: income inequality trends in 16 OECD countries. Am. J. Sociol. 107 (5), 1244–1299.
Alesina, A., Fuchs-Schünde ln, N., 2005. Good Bye Lenin (or Not?): The Effect of Communism on People's Preferences. Working Paper 11700, In: National Bureau of Economic Research (NBER), Cambridge, MA.
Alesina, A., Glaeser, E.L., 2006. Fighting Poverty in the US and Europe: A World of Difference. Oxford University Press, Oxford.
Alesina, A., La Ferrara, E., 2005. Preferences for redistribution in the land of opportunities. J. Public Econ. 89 (5–6), 897–931.
Alesina, A., Giuliano, P., 2009. Preferences for Redistribution. Working Paper 14825, National Bureau of Economic Research (NBER), Cambridge, MA.
Alfaro, L., Kalemli-Ozcan, S., Volosovych, V., 2008. Why doesn't capital flow from rich to poor countries? An empirical investigation. Rev. Econ. Stat. 90 (2), 347–368.
Anand, S., Kanbur, S.M.R., 1993. The Kuznets process and the inequality–development relationship. J. Dev. Econ. 40 (1993), 25–52.
Atif, S.M., Srivastav, M., Sauytbekova, M., Arachchige, U.K., 2012. Globalization and Income Inequality: A Panel Data Analysis of 68 Countries. MPRA Paper No. 42385.
Atkinson, A.B., 1975. The Economics of Inequality. Oxford University Press, Oxford.
Atkinson, A.B., 2002. A critique of the transatlantic consensus on rising income inequality. World Econ. 24 (4), 433–452.
Atkinson, A.B., 2007. The distribution of earnings in OECD countries. Int. Labour Rev. 146 (1–2), 41–61.
Atkinson, A.B., 2008. The Changing Distribution of Earnings in OECD Countries. Oxford University Press, Oxford.
Atkinson, A.B., Bourguignon, F. (Eds.), 2000. Handbook of Income Distribution. Elsevier Science B.V., Amsterdam, p. 958.

Atkinson, A.B., Brandolini, A., 2001. Promise and pitfalls in the use of 'secondary' data-sets: income inequality in OECD countries. J. Econ. Lit. 39 (3), 771–799.
Atkinson, A.B., Brandolini, A., 2006a. On data: a case study of the evolution of income inequality across time and across countries. Camb. J. Econ. 33 (3), 381–404.
Atkinson, A.B., Brandolini, A., 2006b. From earnings dispersion to income inequality. In: Farina, F., Savaglio, E. (Eds.), Inequality and Economic Integration. Routledge, London, pp. 35–64 (chapter 2), 335 p.
Atkinson, A.B., Brandolini, A., 2009. The panel-of-countries approach to explaining income inequality: an interdisciplinary research agenda. In: Morgan, S.L., Grusky, D.B., Fields, G.S. (Eds.), Mobility and Inequality: Frontiers of Research in Sociology and Economics. Stanford University Press, Stanford, pp. 400–448.
Atkinson, A., Marlier, E., 2010. Income and Living Conditions in Europe. Eurostat, Luxembourg.
Atkinson, A.B., Morelli, S., 2014. Chartbook of Economic Inequality, ECINEQ WP 2014 - 324. http://ssrn.com/abstract=2422269.
Atkinson, A.B., Gordon, J.P.F., Harrison, A., 1989. Trends in the shares of top wealth-holders in Britain, 1923–1981. Oxf. Bull. Econ. Stat. 51 (3), 315–332.
Atkinson, A.B., Rainwater, L., Smeeding, T., 1995. Income Distribution in OECD Countries: Evidence from the Luxembourg Income Study. OECD Publishing, Paris.
Atkinson, A.B., Piketty, T., Saez, E., 2011. Top incomes in the long run of history. J. Econ. Lit. 49 (1), 3–71.
Autor, D.H., Katz, L.F., Krueger, A.B., 1998. Computing inequality: have computers changed the labor market? Q. J. Econ. 113 (4), 1169–1213.
Autor, D., Levy, F., Murnane, R., 2003. The skill content of recent technological change: an empirical exploration. Q. J. Econ. 118 (4), 1279–1334.
Babones, S., Alvarez-Rivadulla, M.J., 2007. Standardized income inequality data for use in cross-national research. Socio. Inq. 77 (1), 3–22.
Baccaro, L., 2008. Labour, Globalisation and Inequality: Are Trade Unions Still Redistributive? Discussion Paper No. 192, International Institute for Labour Studies, Geneva.
Ballarino, G., Bratti, M., Filippin, A., Fiorio, C., Leonardi, M., Servini, F., 2014. Increasing educational inequalities? In: Salverda, W. et al., (Ed.), Changing Inequalities in Rich Countries: Analytical and Comparative. Oxford University Press, Oxford (Chapter 5).
Barr, N., 2001. The Welfare State as Piggy Bank. Information, Risk, Uncertainty, and the Role of the State. Oxford University Press, Oxford.
Bassanini, A., Duval, R., 2006. Employment Patterns in OECD Countries: Reassessing the Role of Policies and Institutions. OECD Economics Department Working Papers 486, OECD, Economics Department.
Beck, T., Demirguc-Kunt, A., Levine, R., 2004. Finance, Inequality, and Poverty: Cross-Country Evidence. NBER Working Paper No. 10979, http://www.nber.org/papers/w10979.
Bénabou, R., Ok, E., 2001. Social Mobility and the Demand for Redistribution: The POUM Hypothesis. Q. J. Econ. 116 (2), 447–487.
Beramendi, P., Cusack, T.R., 2009. Diverse disparities: the politics and economics of wage, market, and disposable income inequalities. Polit. Res. Q. 62 (2), 257–275.
Bergh, A., 2005. On the counterfactual problem of welfare state research: how can we measure redistribution? Eur. Sociol. Rev. 21, 345–357.
Bergh, A., Fink, G., 2008. Higher education policy, enrollment, and income inequality. Soc. Sci. Q. 89 (1), 217–235.
Bergh, A., Nilson, Th., 2010. Do liberalization and globalization increase income inequality? Eur. J. Polit. Econ. 26, 488–505.
Berman, E., Bound, J., Machin, S., 1998. Implications of skill-biased technological change: international evidence. Q. J. Econ. 113 (4), 1245–1279.
Bertola, G., 2008. Inequality, Globalization, and Financial Development, Conference Paper Globalization and Inequality: Reflections on the Development of a Divided World, European University Institute in Florence.
Biewen, M., Jenkins, S.P., 2006. Variance estimation for generalized entropy and Atkinson inequality indices: the complex survey data case. Oxf. Bull. Econ. Stat. 68 (3), 371–383.
Birdsall, N., 1998. Life is unfair: inequality in the world. Foreign Policy 112, 76–83.

Blackburn, M.L., Bloom, D., 1987. The Effects of Technological Change on Earnings and Income Inequality in the United States, NBER Working Paper No. 2337, Cambridge.
Blanchard, O., Giavazzi, F., 2003. Macroeconomic effects of regulation and deregulation in goods and labor markets. Q. J. Econ. 118 (3), 879–907, MIT Press.
Blau, F.D., Kahn, L.M., 2009. Inequality and earnings distribution. In: Nolan Salverda, W.B., Smeeding, T.M. (Eds.), pp. 177–203 (Chapter 8).
Bloom, N., Draca, M., Van Reenen, J., 2011. Trade Induced Technical Change: The Impact of Chinese Imports on Innovation and Productivity. NBER Working Paper No. 16717, National Bureau of Economic Research, Cambridge, MA.
Blundell, R.W., 1995. The Impact of Taxation on Labour Force Participation and Labour Supply. OECD Jobs Study Working Papers No. 8, OECD Publishing.
Blundell, R., Bozio, A., Laroque, G., 2011. Labour supply and the extensive margin. Am. Econ. Rev. Pap. Proc. 101 (3), 482–486.
Borck, R., 2007. Voting, inequality and redistribution. J. Econ. Surv. 21 (1), 90–109.
Bourguignon, F., 2005. The effect of economic growth on social structures. In: Aghion, P., Durlauf, S. (Eds.), Handbook of Economic Growth, vol 2. Elsevier, North-Holland, Amsterdam, pp. 1702–1747 (Chapter 27).
Bourguignon, F., Ferreira, F., Lustig, N., 2005. The Microeconomics of Income Distribution Dynamics in East Asia and Latin America. The World Bank and Oxford University Press, Washington, DC, 2005.
Bourguignon, F., Bussolo, M., Cockburn, J., 2010. Macro–micro analytics: background, motivation, advantages and remaining challenges. Int. J. Microsimulation 3 (1), 1–7.
Bradley, D., Huber, E., Moller, S., Nielsen, F., Stephens, J., 2003. Distribution and redistribution in postindustrial democracies. World Polit. 55 (January), 193–228.
Brady, D., 2006. Structural theory and relative poverty in rich western democracies, 1969–2000. Res. Soc. Stratif. Mobil. 24 (2), 153–175.
Brandolini, A., 2003. A bird-eye view of long-run changes in income inequality. Bank of Italy, mimeo.
Brandolini, A., Smeeding, T.M., 2009. Income inequality in richer and OECD countries. In: Salverda, W., Nolan, B., Smeeding, T.M. (Eds.), The Oxford Handbook of Economic Inequality, Oxford University Press, Oxford, pp. 71–100.
Bruno, G.S.F., Crino, R., Falzoni, A.M., 2004. Foreign Direct Investment, Wage Inequality, and Skilled Labor Demand in EU Accession Countries, Centro Studi Luca d'Agliano Development Studies Working Papers No. 188, October 2004.
Burniaux, J.M., Padrini, F., Brandt, N., 2006. Labour Market Performance, Income Inequality and Poverty in OECD Countries, OECD Economics Department Working Papers 500.
Burtless, G., 1999. Effects of growing wage disparities and changing family composition on the US income distribution. Eur. Econ. Rev. 43, 853–865.
Burtless, G., 2009. Demographic transformation and economic inequality. In: Salverda, W., Nolan, B., Smeeding, T.M. (Eds.), The Oxford Handbook of Economic Inequality, Oxford University Press, Oxford, pp. 435–454 (Chapter 18).
Card, D., 1996. The effect of unions on the structure of wages: a longitudinal analysis. Econometrica 64 (4), 957–979.
Card, D., DiNardo, J., 2002. Skill Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles, NBER Working Paper No. 8769.
Carnoy, M., 2011. As higher education expands, is it contributing to greater inequality? Natl. Inst. Econ. Rev. 215, R4.
Carter, J.R., 2007. An empirical note on economic freedom and income inequality. Public Choice 130 (1–2), 163–177.
Cassette, A., Fleury, N., Petit, S., 2012. Income inequalities and international trade in goods and services: short- and long-run evidence. Int. Trade J. 26 (3), 223–254.
Çelik, S., Basdas, U., 2010. How does globalization affect income inequality? A panel data analysis. Int. Adv. Econ. Res. 16 (4), 358–370. Accessed June 19, 2012, http://www.springerlink.com/index/10.1007/s11294-010-9281-0.
Checchi, D., Garcia-Penalosa, C., 2005. Labour Market Institutions and the Personal Distribution of Income in the OECD, IZA Discussion Paper 1681.

Checchi, D., Garcia-Penalosa, C., 2008. Labour market institutions and income inequality. Econ. Policy 23 (56), 601–649, CEPR, CES, MSH.
Chen, 2013. Assessing the Possible Distributional Impact of Migration in OECD Countries—A Scoping Paper, Discussion Paper at the EU/OECD Seminar on Jobs, Wages and Inequality, Paris November 2013, mimeo.
Chen, W.-H., Förster, M., Llena-Nozal, A., 2013a. Globalisation, Technological Progress and Changes in Regulations and Institutions—Which Impact on the Rise of Earnings Inequality in OECD Countries? Luxembourg Income Study Working Paper, No. 597.
Chen, W.-H., Förster, M., Llena-Nozal, A., 2013b. Determinants of Households Earnings Inequality: The Role of Labour Market Trends and Changing Household Structure, Luxembourg Income Study Working Paper, No. 591.
Chen, W.-H., Förster, M., Marshalian, M., Llena-Nozal, A., 2014. Women, Work and Wages: The Gender Aspect of Earnings and Household Income Inequality. OECD Social, Employment and Migration Working Papers, OECD Publishing, Paris (forthcoming).
Chusseau, N., Dumont, 2012. Growing Income Inequalities in Advanced Countries, ECINEQ Working Paper No. 2012-260.
Chusseau, N., Dumont, M., Hellier, J., 2008. Explaining rising inequality: skill-biased technical change and north–south trade. J. Econ. Surv. 22 (3), 409–457.
Corneo, G., Grüner, H.P., 2002. Individual preferences for political redistribution. J. Public Econ. 83, 83–107.
Cornia, G.A., 2005. Policy Reform and Income Distribution, DESA Working Paper No. 3.
Cornia, G.A., 2012. Inequality Trends and their Determinants: Latin America Over 1990–2010. WIDER Working Paper No. 2012/09 appeared under the same title as Chapter 2 in Cornia, G.A. (Ed.), 2014. Falling Inequality in Latin America. Oxford University Press, pp. 23–49.
Cusack, T., Fuchs, S., 2002. Ideology, Institutions, and Public Spending. Discussion Paper P 02 – 903, Social Science Research Center, Berlin.
Davidson, R., Flachaire, E., 2007. Asymptotic and bootstrap inference for inequality and poverty measures. J. Econom. 141 (1), 141–166.
Davis, D., Mishra, P., 2007. Stolper–Samuelson is dead: and other crimes of both theory and data. In: Harrison, A. (Ed.), Globalization and Poverty, University of Chicago Press, Chicago, pp. 87–107.
De Gregorio, J., Lee, J., 2002. Education and income inequality: new evidence from cross country data. Rev. Income Wealth 48 (3), 395–417.
Deininger, K., Squire, -L., 1996. A new data set measuring income inequality. World Bank Econ. Rev. 10 (3), 565–591.
Deininger, K., Squire, L., 1998. New ways of looking at old issues. J. Dev. Econ. 57, 259–287.
Dickens, R., Machin, S., Manning, A., 1999. The effects of minimum wages on employment: theory and evidence from Britain. J. Labor Econ. 17, 1–22.
DiNardo, J., Fortin, N., Lemieux, T., 1996. Labour market institutions and the distribution of wages, 1973–1992: a semi-parametric approach. Econometrica 64, 1001–1044.
DiPrete, Th., McManus, P., 2000. Family change, employment transitions, and the welfare state: household income dynamics in the United States and Germany. Am. Sociol. Rev. 65 (3), 343–370.
Doerrenberg, P., Peichl, A., 2012. The Impact of Redistributive Policies on Inequality in OECD Countries, IZA DP No. 6505.
Dollar, D., Kraay, A., 2002. Growth is good for the poor. J. Econ. Growth 7, 195–225.
Downs, A., 1957. An Economic Theory of Democracy. Harper and Brothers, New York.
Dreher, A., 2006. Does globalization affect growth? Evidence from a new index of globalization. Appl. Econ. 38 (10), 1091–1110.
Dreher, A., Gaston, N., 2008. Has globalization increased inequality? Rev. Int. Econ. 16 (3), 516–536.
Ebenstein, A., Harrison, A., McMillan, M., Phillips, S., 2009. Estimating the Impact of Trade and Offshoring on American Workers Using the Current Population Surveys, NBER Working Paper No. 15107.
Eberhardt, M., Teal, F., 2009. Econometrics for Grumblers: A New Look at the Literature on Cross-Country Growth Empirics, CSAE WPS/2009-07.

Egger, H., Kreickemeier, U., 2009. Firm heterogeneity and the labor market effects of trade liberalization. Int. Econ. Rev. 50 (1), 187–216.
Egger, H., Kreickemeier, U., 2010. Worker-specific effects of globalisation. World Econ. 33 (8), 987–1005.
Esping-Andersen, G., 1990. The Three Worlds of Welfare Capitalism. Princeton University Press, Princeton, NJ.
Esping-Andersen, G., 2009. The Incomplete Revolution: Adapting to Women's New Roles. Polity Press, Cambridge, UK.
Esping-Andersen, G., Myles, J., 2009. Economic inequality and the welfare state. In: Salverda, W., Nolan, B., Smeeding, T. (Eds.), The Oxford Handbook of Economic Inequality. Oxford University Press, Oxford, pp. 639–664.
Faustino, H., Vali, C., 2012. The Effects of Globalisation on OECD Income Inequality: A Static and Dynamic Analysis, School of Economics and Management Working Paper WP 12/2011/DE, Lisbon.
Feenstra, R., Hanson, G., 1996. Globalization, outsourcing, and wage inequality. Am. Econ. Rev. 86, 240–245.
Feenstra, R., Hanson, G.H., 1999. The impact of outsourcing and high-technology capital on wages: estimates for the United States, 1979–1990. Q. J. Econ. 114 (3), 907–940.
Feenstra, R., Hanson, G., 2003. Global production sharing and rising inequality: a survey of trade and wage. In: Choi, E.K., Harrigan, J. (Eds.), Handbook of International Trade. Blackwell, Malden, MA, pp. 146–185.
Ferreira, F.H.G., Ravallion, M., 2009. Poverty and inequality: the global context. In: Salverda, W., Nolan, B., Smeeding, T.M. (Eds.), Chapter 24, pp. 599–636.
Figini, P., Görg, H., 2006. Does Foreign Direct Investment Affect Wage Inequality? An empirical Investigation, IZA Discussion Paper No. 2336.
Finseraas, H., 2008. Income inequality and demand for redistribution: a multilevel analysis of european public opinion. Scand. Polit. Stud. Nord. Polit. Sci. Assoc. 32 (1), 94–119.
Fiori, G., Nicoletti, G., Scarpetta, S., Schiantarelli, F., 2007. Employment Outcomes and the Interaction between Product and Labor Market Deregulation: Are They Substitutes or Complements? IZA Discussion Papers 2770, Institute for the Study of Labor (IZA).
Firpo, S., Fortin, N.M., Lemieux, T., 2009. Unconditional quantile regressions. Econometrica 77 (3), 953–973.
Fong, C.M., 2001. Social preferences, self-interest, and the demand for redistribution. J. Public Econ. 82 (2), 225–246.
Fong, C.M., 2006. Prospective Mobility, Fairness, and the Demand for Redistribution. Department of Social and Decision Sciences Working Paper, Carnegie Mellon University, Pittsburgh, PA.
Förster, M., 2013. Répartition des revenus et inégalités sociales. In: Bontout, O., Hazouard, S., Lasserre, R., Zaidman, C. (Eds.), Les Réformes de la Protection Sociale en Allemagne. CIRAC/DREES, Paris, pp. 237–258.
Förster, M.F., Mira d'Ercole, M., 2012. The OECD approach to measuring income distribution and poverty. In: Besharov, D., Couch, K. (Eds.), Counting the Poor—New Thinking About European Poverty Measures and Lessons for the United States. Oxford University Press, Oxford, pp. 27–58.
Förster, M.F., Verbist, G., 2012. Money or Kindergarten? Distributive Effects of Cash Versus In-Kind Family Transfers for Young Children. OECD Social, Employment and Migration Working Papers No. 135, OECD Publishing.
Förster, M.F., Whiteford, P., 2009. How much redistribution do Welfare States achieve? The role of cash transfers and household taxes. CESifo DICE REPORT J. Inst. Comp. 7 (3), 34–41, Munich.
Francois, J., Nelson, D., 2003. Globalization and Relative Wages: Some Theory and Evidence. GEP Research Paper 03/15, University of Nottingham.
Francois, J., Rojas-Romagosa, H., 2005. The Construction and Interpretation of Combined Cross-Section and Time-Series Inequality Datasets. CEPR Discussion Paper No. 5214.

Freeman, R., 1993. How much has de-unionization contributed to the rise in male earnings inequality?" In: Danziger, S., Gottschalk, P. (Eds.), Uneven Tides: Rising Inequality in America. Russell Sage, New York, NY, pp. 133–163.

Freeman, R., 2009. Globalization and inequality. In: Salverda, W., Nolan, B., Smeeding, T. (Eds.), The Oxford Handbook of Economic Inequality. Oxford University Press, Oxford, pp. 575–589.

Fuest, C., Niehues, J., Peichl, A., 2009. The Redistributive Effects of Tax Benefit Systems in the Enlarged EU. IZA DP No. 4520.

Galbraith, J., 2012. Inequality and Instability. A Study of the World Economy Just Before the Crisis. Oxford University Press, Oxford.

Gijsberts, M., 2002. The legitimation of income inequality in state-socialist and market societies. Acta Sociol. 45 (4), 269–285.

Goedemé, T., 2013. How much confidence can we have in EUSILC? Complex sample designs and the standard error of the Europe 2020 poverty indicators. Soc. Indic. Res. 110 (1), 89–110. http://dx.doi.org/10.1007/s11205-011-9918-2.

Goldberg, P.K., Pavcnik, N., 2007. Distributional effects of globalisation in developing countries. J. Econ. Lit. XLV, 39–82.

Golden, M.A., Wallerstein, M., 2011. Domestic and international causes for the rise of pay inequality in OECD nations between 1980 and 2000. In: Brady, D. (Ed.), Comparing European Workers Part A (Research in the Sociology of Work, Volume 22 Part 1) Emerald Group Publishing Limited, pp. 209–249.

Goos, M., Manning, A., 2007. Lousy and lovely jobs: the rising polarization of work in Britain. Rev. Econ. Stat. 89 (1), 118–133, 01.

Gottschalk, P., Smeeding, T.M., 2000. Empirical evidence on income inequality in industrial countries. In: Atkinson, A.B., Bourguignon, F. (Eds.), Handbook of Incomes Distribution. Elsevier, North-Holland, Amsterdam, pp. 261–307 (Chapter 5).

Goudswaard, K., Caminada, K., 2010. The redistributive effect of public and private social programmes: a cross-country empirical analysis. Int. Soc. Secur. Rev. 63 (1), 1–19.

Gourdon, J., Maystre, N., de Melo, J., 2008. Openness, inequality and poverty: endowments matter. J. Int. Trade Econ. Dev. 17 (3), 343–378.

Greenwood, J., Guner, N., Kocharkov, G., Santos, C., 2014. Marry Your Like: Assortative Mating and Income Inequality, NBER Working Paper No. 19829.

Guillaud, E., 2013. Preferences for redistribution: an empirical analysis over 33 countries. J. Econ. Inequal. 11 (1), 57–78.

Hall, P.A., Soskice, D. (Eds.), 2001. Varieties of Capitalism. The Institutional Foundations of Comparative Advantage, Oxford University Press, Oxford.

Haskel, J.E., Slaughter, M., 1999. Trade, Technology and UK Wage Inequality? CEPR Discussion Paper No. 2091.

Haskel, J.E., Slaughter, M., 2001. Trade, technology and UK wage inequality. Econ. J. 111, 163–187.

Haskel, J.E., Slaughter, M., 2002. Does the sector bias of skill-biased technical change explain changing skill premia? Eur. Econ. Rev. 46, 1757–1783.

Hellier, J., Lambrecht, S., 2012. Inequality, growth and welfare: the main links. In: Hellier, J., Chusseau, N. (Eds.), Growing Income Inequalities: Economic Analyses. Palgrave Macmillan, pp. 274–311, http://www.palgrave.com/page/detail/?sf1=id_product&st1=520792&loc=uk.

Helpman, E., Itskhoki, O., Redding, S., 2009. Inequality and Unemployment in a Global Economy, NBER Working Paper No. 14478, Cambridge, MA.

Heshmati, A., 2004. The Relationship Between Income Inequality, Poverty and Globalisation, IZA Discussion Paper No. 1277.

Hijzen, A., 2007. International outsourcing, technological change, and wage inequality. Rev. Int. Econ. 15 (1), 188–205.

Hijzen, A., Martins, P.S., Schank, T., Upward, R., 2013. Foreign-owned firms around the world: a comparative analysis of wages and employment at the micro-level. European Economic Review 60, 170–188.

Hills, J., 2004. Inequality and the State. Oxford University Press, Oxford.

Hox, J., 2002. Multilevel Analyses. Lawrence Erlbaum Assoc, Mahwaj, NJ.
Iacovou, M., Levy, H., Kaminska, O., 2012. Using EU-SILC Data for Cross-National Analysis: Strengths, Problems and Recommendations. ISER Working Paper No. 2012-03, University of Essex.
ILO, 2008. Labour institutions and inequality. In: ILO, World of Work Report, Volume October 2008. ILO. Geneva, pp. 71–114 (Chapter 3).
IMF, 2007. Globalization and inequality. In: World Economic Outlook, Volume October 2007. IMF, Washington, pp. 31–65.
Immervoll, H., Richardson, L., 2011. Redistribution Policy and Inequality Reduction in OECD Countries: What Has Changed in Two Decades? LIS Working Paper No. 571.
Iversen, T., Soskice, D., 2006. Electoral institutions and the politics of coalitions: why some democracies redistribute more than others. Am. Polit. Sci. Rev. 100, 165–181.
Iversen, T., Soskice, D., 2009. Distribution and redistribution. The shadow of the nineteenth century. World Polit. 61 (3), 438–486.
Jain, S., 1975. Size Distribution of Income. World Bank, Washington.
Janmaat, J.G., 2013. Subjective inequality: a review of international comparative studies on people's views about inequality. Eur. J. Sociol. 54 (03), 357–389. http://dx.doi.org/10.1017/S000397561300020.
Jäntti, M., 1997. The role of demographic shifts, markets, and government policies. Economica 64 (255), 415–440.
Jäntti, M., Jenkins, S., 2010. The impact of macroeconomic conditions on income inequality. J. Econ. Inequal. 8 (2), 221–240, Springer.
Jaumotte, F., Lall, S., Papageorgiou, C., 2008. Rising Income Inequality: Technology, or Trade and Financial Globalization? IMF Working Paper No. WP/08/185.
Jesuit, D.K., Mahler, V.A., 2010. Comparing government redistribution across countries: the problem of second order effects. Soc. Sci. Q. 91 (5), 1390–1404.
Kahanec, M., Zimmermann, K., 2009. International migration, ethnicity, and economic inequality. In: Salverda, W., Nolan, B., Smeeding, T. (Eds.), The Oxford Handbook of Economic Inequality. Oxford University Press, Oxford, pp. 455–490.
Kammer, A., Niehues, J., Peichl, A., 2012. Welfare regimes and welfare state outcomes in Europe. J. Eur. Soc. Policy 22, 455–471.
Kang, B., Yun, M., 2008. Changes in Korean Wage Inequality, 1980–2005, IZA Discussion Paper No. 3780.
Karoly, L., Burtless, G., 1995. Demographic change, rising earnings inequality, and the distribution of personal well-being, 1959–1989. Demography 32 (3), 379–405.
Kearney, A.T., Inc., the Carnegie Endowment for International Peace, 2004. Measuring globalization: economic reversals, forward momentum. Foreign Pol. March/April 2004, 1–12.
Kearney, A.T., Inc., the Carnegie Endowment for International Peace, 2007. The globalization index 2007. Foreign Pol. October/November 2007, 1–12.
Keely, L.C., Tan, C.M., 2008. Understanding preferences for income redistribution. J. Public Econ. 92 (5–6), 944–961.
Kelley, J., Zagorski, K., 2004. Economic change and the legitimation of inequality: the transition from socialism to the free market in Central-East Europe. In: Bills, D.B. (Ed.), Research in Social Stratification and Mobility. vol. 22, Elsevier Ltd, Oxford, pp. 319–364.
Kenworthy, L., 2007. Inequality and sociology. Am. Behav. Sci. 50 (5), 584–602. http://dx.doi.org/10.1177/0002764206295008.
Kenworthy, L., 2011. Progress for the Poor. Oxford University Press, Oxford.
Kenworthy, L., McCall, L., 2007. Inequality, Public Opinion, and Redistribution. Working Paper 459, LIS, Luxembourg.
Kenworthy, L., Pontusson, J., 2005. Rising inequality and the politics of redistribution in affluent countries. Perspect. Polit. 3 (3), 449–471.
Koeninger, W., Leonardi, M., Nunziata, L., 2007. Labour market institutions and wage inequality. Ind. Labour Relat. Rev. 60 (3), 340–356.
Korpi, W., 2006. Power resources and employer-centered approaches in explanations of welfare states and varieties of capitalism: protagonists, consenters, and antagonists. World Polit. 58 (2), 167–206.

Korpi, W., Palme, J., 1998. The paradox of redistribution and strategies of equality. Am. Sociol. Rev. 63, 661–687.
Korpi, W., Palme, J., 2003. New politics and class politics in the context of austerity and globalization: welfare state regress in 18 countries, 1975–95. Am. Polit. Sci. Rev. 97 (3), 425–446.
Kovacevic, M.S., Binder, D.A., 1997. Variance estimation for measures of income inequality and polarization—the estimating equations approach. J. Off. Stat. 13 (1), 41–58.
Kremer, M., Maskin, E., 2003. Globalization and Inequality, Harvard University, Department of Economics. Unpublished Manuscript.
Kristov, L., Lindert, P., McLelland, R., 1992. Pressure groups and redistribution. J. Public Econ. 48, 135–163.
Krueger, A.B., 1993. How computers have changed the wage structure: evidence from microdata, 1984–1989. Q. J. Econ. 108, 33–60.
Krugman, P., 1995. Growing world trade: causes and consequences. Brookings Pap. Econ. Act. 1, 327–377.
Krugman, P., 2000. Technology, trade and factor prices. J. Int. Econ. 50, 51–71.
Krugman, P., 2007. Trade and Inequality, Revisited. http://voxeu.org/index.php?q=node/261.
Kuznets, S., 1955. Economic growth and income inequality. Am. Econ. Rev. 65, 1–25.
Lambert, P., Nesbakken, R., Thoresen, Th., 2010. On the Meaning and Measurement of Redistribution in Cross-Country Comparisons, LIS Working Paper No. 532.
Lanz, R., Miroudot, S., Nordås, H.K., 2011. Trade in Tasks. OECD Trade Policy Working Papers No. 117, OECD Publishing.
Larcinese, V., 2007. Voting over redistribution and the size of the welfare state: the role of turnout. Polit. Stud. 55 (3), 568–585.
Layard, R., Nickell, S., Jackman, R., 1991. Unemployment: Macroeconomic Performance and the Labour Market. Oxford University Press, Oxford.
Lee, D.S., 1999. Wage inequality during the 1980s: rising dispersion or falling minimum wage? Q. J. Econ. 114, 977–1023.
Lee, E., Vivarelli, M., 2006. Understanding Globalization, Employment and Poverty Reduction. Palgrave, Houndmills.
Lemieux, T., 2008. The changing nature of wage inequality. J. Popul. Econ. 21 (1), 21–48.
Li, H., Zou, H., Squire, 1998. Explaining international and intertemporal variations in income inequality. Econ. J. 108, 26–43. http://www.worldbank.org/research/inequality/pdf/squire.pdf.
Lorentowicz, A., Marin, D., Raubold, A., 2005. Is Human Capital Losing from Outsourcing? Evidence from Austria and Poland, Governance and the Efficiency of Economic Systems, Discussion Paper No. 76.
Lu, Y., Morissette, R., Schirle, T., 2011. The growth of family earnings inequality in Canada, 1980–2005. Rev. Income Wealth 57 (1), 23–39.
Lübker, M., 2007. Inequality and the demand for redistribution: are the assumptions of the new growth theory valid? Soc Econ Rev. 5 (1), 117–148.
Lucas, R., 1990. Why doesn't capital flow from rich to poor countries? Am. Econ. Rev. 80 (2), 92–96.
Lupu, N., Pontusson, J., 2011. The structure of inequality and the politics of redistribution. Am. Polit. Sci. Rev. 105 (2), 316–336.
Luttmer, E., Singhal, M., 2008. Culture, Context, and the Taste for Redistribution. Working Paper 14268, National Bureau of Economic Research (NBER), Cambridge, MA.
Machin, S., 1997. The decline of labour market institutions and the rise of wage inequality in Britain. Eur. Econ. Rev. 41 (3–5), 647–657.
Machin, S., 2009. Education and inequality. In: Salverda, W., Nolan, B., Smeeding, T.M. (Eds.), The Oxford Handbook of Economic Inequality. Oxford University Press, Oxford, pp. 261–307 (Chapter 17).
Machin, S., Van Reenen, J., 1998. Technology and Changes in Skill Structure: Evidence from Seven OECD Countries. IFS Working Paper W98/04, Institute for Fiscal Studies.
Mahler, V.A., 2004. Economic globalization, domestic politics, and income inequality in the developed countries: a cross-national study. Comp. Polit. Stud. 37, 1025–1053. http://cps.sagepub.com/cgi/content/abstract/37/9/10257.

Mahler, V.A., 2008. Electoral turnout and income redistribution by the state: a cross-national analysis of the developed democracies. Eur. J. Polit. Res. 47 (2), 161–183.
Mahler, V.A., 2010. Government inequality reduction in comparative perspective: a cross-national study of the developed world. Polity. 42 (4), 511(31) (Report).
Mahler, V., Jesuit, D., 2013. Political sources of government redistribution in high-income countries. In: Gornick, J.C., Jäntti, M. (Eds.), Economic Inequality in Cross-National Perspective. Stanford University Press, Stanford, pp. 145–172.
Mankiw, G., 1995. The growth of nations. Brookings Pap. Econ. Act. 1995 (1), 275–326, 25th Anniversary Issue (1995).
Marx, I., Salanauskaite, L., Verbist, G., 2013. The Paradox of Redistribution Revisited: And that it May Rest in Peace? IZA Discussion Paper 7414 (May 2013). Available at http://www.iza.org/en/webcontent/publications/papers/viewAbstract?dp_id=7414.
McCarty, N., Pontusson, J., 2009. The political economy of inequality and redistribution. In: Salverda, W., Nolan, B., Smeeding, T.M. (Eds.), The Oxford Handbook of Economic Inequality. Oxford University Press, Oxford, pp. 665–692.
Medgyesi, M., Tóth, I.Gy., 2009. Economic growth and income inequalities. In: Ward, T., Lelkes, O., Sutherland, H., Tóth, I.Gy. (Eds.), European Inequalities. Tárki Social Research Institute, Budapest, pp. 131–152 (Chapter 6).
Melitz, M., 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. Econometrica 71 (5), 1695–1725.
Meltzer, A.H., Richard, S.F., 1981. A rational theory of the size of government. J. Polit. Econ. 89 (5), 914–927.
Messina, J., 2003. The Role of Product Market Regulations in the Process of Structural Change. Working Paper Series, No. 217, European Central Bank.
Michaels, G., Natraj, A., Van Reenen, J., 2010. Has ICT Polarized Skill Demand? Evidence from Eleven Countries over 25 Years, CEPR Discussion Papers 7898, CEPR Discussion Papers.
Milanovic, B., 2000. The median-voter hypothesis, income inequality and income redistribution. Eur. J. Polit. Econ. 16 (2–3), 367–410.
Milanovic, B., 2005. Can we discern the effect of globalization on income distribution? Evidence from household surveys. World Bank Econ. Rev. 191, 21–44.
Milanovic, B. (Ed.), 2012. Globalization and Inequality, Edward Elgar Publishing, Cheltenham, UK.
Milanovic, B., Squire, I., 2007. Does tariff liberalization increase wage inequality. In: Harrison, A. (Ed.), Globalization and Poverty. Chicago University Press, Chicago, pp. 143–181 (chapter 4).
Moene, K.O., Wallerstein, M., 2001. Inequality, social insurance, and redistribution. Am. Polit. Sci. Rev. 95 (4), 859–874.
Moene, K.O., Wallerstein, M., 2003. Earnings inequality and welfare spending: a disaggregated analysis. World Polit. 55 (4), 485–516.
Mohl, P., Pamp, O., 2009. Income inequality & redistributional spending: an empirical investigation of competing theories. Public Finance Manag. 9 (2), 179(56) (Report).
Muinelo-Gallo, l., Roca-Sagalés, O., 2011. Economic growth and inequality: the role of fiscal policies. Aust. Econ. Pap. 50 (2–3), 74–97.
Muinelo-Gallo, l., Roca-Sagalés, O., 2013. Joint determinants of fiscal policy, income inequality and economic growth. Econ. Model. 30, 814–824.
Nel, P., 2005. Democratization and the dynamics of income distribution in low- and middle-income countries. Politikon 32 (1), 17–43. http://dx.doi.org/10.1080/02589340500101675.
Nickell, S., 1998. Unemployment: questions and some answers. Econ. J. Roy. Econ. Soc. 108 (448), 802–816.
Nicoletti, G., Scarpetta, S., 2005. Product Market Reforms and Employment in OECD Countries. OECD Economics Department Working Papers 472OECD, Economics Department.
Nicoletti, G., Haffner, R., Nickell, S., Scarpetta, S., Zoega, G., 2001. European integration, liberalization and labor market performance. In: Bertola, G., Boeri, T., Nicoletti, G. (Eds.), Welfare and Employment in United Europe. MIT Press, Cambridge, MA, pp. 147–236.
Niehues, J., 2010. Social Spending Generosity and Income Inequality: A Dynamic Panel Approach, IZA Discussion Paper Series. Discussion Paper No. 5178.

Nielsen, F., Alderson, A., 1995. Income inequality, development, and dualism: results from an unbalanced cross-national panel. Am. Sociol. Rev. 60, 674–701.
Nolan, B., Marx, I., 2009. Economic inequality, poverty, and social exclusion. In: Salverda, W., Nolan, B., Smeeding, T.M. (Eds.), The Oxford Handbook of Economic Inequality. Oxford University Press, Oxford, pp. 315–341 (Chapter 13).
Nolan, B., Salverda, W., Checchi, D., Marx, I., Mcknight, A., Tóth, I.Gy., van de Werfhorst, H.G. (Eds.), 2014. Changing Inequalities and Societal Impacts in Rich Countries: Thirty Countries' Experiences. Oxford University Press, Oxford.
Nollmann, G., 2006. Erhöht Globalisierung die Ungleichheit der Einkommen? Determinanten von Einkommensverteilungen in 16 OECD-Ländern 1967–2000. Kölner Zeitschrift für Soziologie und Sozialpsychologie 58 (4), 638–659.
Nunziata, L., 2002. Unemployment, Labour Market Institutions and Shocks. Economics Papers 2002-W16, Economics Group, Nuffield College, University of Oxford.
OECD, 2005. Measuring Globalisation. OECD Economic Globalisation Indicators. OECD Publishing, Paris.
OECD, 2006. OECD Employment Outlook - Boosting Jobs and Incomes. OECD Publishing, Paris.
OECD, 2007. Offshoring and Employment: Trends and Impacts. OECD Publishing, Paris.
OECD, 2008. Growing Unequal? Income Distribution and Poverty in OECD Countries. OECD Publishing, Paris, 308 pp.
OECD, 2011. Divided We Stand: Why Inequality Keeps Rising? OECD Publishing, Paris, 386 pp.
OECD, 2013a. Crisis Squeezes Income and Puts Pressure on Inequality and poverty—New Results from the OECD Income Distribution Database. OECD Publishing, Paris.
OECD, 2013b. Quality Review of the OECD Database of Household Incomes and Poverty and the OECD Earnings Database. OECD Publishing, Paris.
Olivera, J., 2014. Preferences for redistribution after the economic crisis. Working Papers, ECINEQ, Society for the Study of Economic Inequality 334, ECINEQ. Society for the Study of Economic Inequality.
Olson, M., 1965. The Logic of Collective Action Cambridge. Cambridge University Press, Cambridge, UK.
Osberg, L., Smeeding, T.M., 2006. 'Fair' inequality? Attitudes toward pay differentials: the united states in comparative perspective. Am. Sociol. Rev. 71 (3), 450–473.
Osberg, L., Smeeding, T.M., Schwabish, J., 2004. Income distribution and public social expenditure: theories, effects, and evidence. In: Neckerman, K. (Ed.), Social Inequality. Russel Sage Foundation, New York, NY, pp. 821–859.
Osier, G., Berger, Y., Goedemé, T., 2013. Standard error estimation for the EU-SILC indicators of poverty and social exclusion. Eurostat Methodologies and Working Papers, Eurostat, Luxembourg, 2013 edition.
Palma, J.G., 2011. Homogeneous middles vs. heterogeneous tails, and the end of the 'inverted-U': it's all about the share of the rich. Dev. Change 42 (1), 87–153.
Parker, S.C., 2000. Opening a can of worms: the pitfalls of time-series regression analyses of income inequality. Appl. Econ. 32 (2), 221–230.
Paukert, F., 1973. Income distribution at different levels of development: a survey of evidence. Int. Labour Rev. 108, 97–125.
Pavcnik, N., 2011. Globalization and within-country income inequality. In: Bacchetta, M., Jansen, M. (Eds.), Making Globalization Socially Sustainable, International Labour Organization and World Trade Organization publication, pp. 233–260 (Chapter 7).
Peichl, A., Pestel, N., Schneider, H., 2010. Does Size Matter? The Impact of Changes in Household Structure on Income Distribution in Germany, IZA DP No. 4770.
Piketty, T., 1995. Social mobility and redistributive politics. Q. J. Econ. 110 (3), 551–584.
Piketty, T., Saez, E., 2006. The evolution of top incomes: a historical and international perspectives. Am. Econ. Rev. 96 (2), 200–205.
Pissarides, C., 1990. Equilibrium Unemployment Theory. Basil Blackwell, Oxford.
Pontusson, J., 2000. Wage inequality and varieties of capitalism. World Polit. 52, 350–383.
Pontusson, J., Rueda, D., 2010. The politics of inequality: voter mobilization and left parties in advanced industrial states. Comp. Polit. Stud. 43 (6), 675–705. http://dx.doi.org/10.1177/0010414009358672.

Pontusson, J., Rueda, D., Way, C.R., 2002. Comparative political economy of wage distribution: the role of partisanship and labour market institutions. Br. J. Polit. Sci. 32 (2), 281–308. http://dx.doi.org/10.1017/S000712340200011X.

Przeworski, A., Teune, H., 1970. The Logic of Comparative Social Inquiry. Wiley-Interscience, New York, 153 p.

Pyatt, G., 2003. Development and the distribution of living standards: a critique of the evolving data base. Rev. Income Wealth 49 (3), 333–358.

Ravallion, M., Loskhin, M., 2000. Who wants to redistribute? The tunnel effect in 1990s Russia. J. Public Econ. 76, 87–104.

Rehme, G., 2007. Education, economic growth and measured income inequality. Economica 74 (295), 493–514.

Reuveny, R., Li, Q., 2003. Economic openness, democracy, and income inequality: an empirical analysis. Comp. Polit. Stud. 36 (5), 575–601.

Robinson, J.A., 2009. The Political Economy of Redistributive Policies, Research for Public Policy, Inclusive Development, ID-09–2009. RBLAC-UNDP, New York.

Roemer, J.E., 1998. Why the poor do not expropriate the rich: an old argument in new grab. J. Public Econ. 70, 399–422.

Rohrbach, D., 2009. Sector bias and sector dualism: the knowledge society and inequality. Int. J. Comp. Sociol. 50 (5–6), 510–536.

Romer, T., 1975. Individual welfare, majority voting, and the properties of a linear income tax. J. Public Econ. 14 (May), 163–185.

Rueda, D., 2008. Left government, policy, and corporatism: explaining the influence of partisanship on inequality. World Polit. 60 (3), 349–389.

Rueda, D., Pontusson, J., 2000. Wage inequality and varieties of capitalism. World Polit. 52, 350–383.

Sabirianova-Peter, K., Buttrick, P., Duncan, D., 2010. Global reform of personal income taxation, 1981–2005: Evidence from 189 countries. Natl. Tax J. 63 (3), 447–478.

Salverda, W., Nolan, B., Smeeding, T. (Eds.), 2009. The Oxford Handbook of Economic Inequality, Oxford University Press, Oxford.

Salverda, W., Nolan, B., Checchi, D., Marx, I., Mcknight, A., Tóth, I.Gy., van de Werfhorst, H.G. (Eds.), 2014. Changing Inequalities and Societal Impacts in Rich Countries: Analytical and Comparative Perspectives, Oxford University Press, Oxford.

Scheve, K.F., Slaughter, M.J., 2007. A new deal for globalization. Foreign Aff. 86 (4), 34–47.

Schwartz, C., 2010. Earnings inequality and the changing association between spouses' earnings. Am. J. Sociol. 115 (5), 1524–1557.

Senik, C., 2009. Income Distribution and Subjective Happiness: A Survey. OECD Social, Employment and Migration Working Paper 96, OECD, Paris, France.

Shorrocks, A., 1982. Inequality decomposition by factor components. Econometrica 50, 193–211.

Shorrocks, A., Guanghua, W., 2008. Ungrouping Income Distributions - Synthesising Samples for Inequality and Poverty Analysis. WIDER Research Paper No. 2008/16.

Slaughter, M.J., 2000. Production transfer within multinational enterprises and American wages. J. Int. Econ. 50, 449–472.

Smeeding, T., 2002. Globalization, inequality, and the rich countries of the G-20: evidence from the Luxembourg Income Study (LIS). In: Gruen, J.L.D., OBrien, T. (Eds.), Globalisation, Living Standards, and Inequality, Recent Progress and Continuing Challenges. J.S. McMillian Printing Group, Lidcombe NSW, Australia, pp. 179–206.

Solt, F., 2009. Standardizing the world income inequality database. Soc. Sci. Q. 90 (2), 231–242.

Spector, D., 2004. Competition and the capital-labor conflict. Eur. Econ. Rev. 48 (1), 25–38, Elsevier.

Spilimbergo, A., Londoño, J.L., Székely, M., 1999. Income distribution, factor endowments, and trade openness. J. Dev. Econ. 59 (1), 77–101.

Ståhlberg, A.C., 2007. Redistribution across the life course in social protection systems. In: OECD (Eds.), Modernising Social Policy for the New Life Course. OECD Publishing, Paris, pp. 201–216, 226 p.

Suhrcke, M., 2001. Preferences for Inequality. East vs. West. Innocenti Working Paper 89, UNICEF International Child Development Centre, Florence, Italy.
Sylwester, K., 2003. Enrolment in higher education and changes in income inequality. Bull. Econ. Res. 55, 249–262.
Taylor, K., Driffield, N., 2005. Wage inequality and the role of multinationals: evidence from UK panel data. Labour Econ. 12, 223–249.
Te Velde, D., Xenogiani, T., 2007. Foreign direct investment and international skill inequality. Oxf. Dev. Stud. Taylor Francis J. 35 (1), 83–104.
Temple, J., 2000. Growth regressions and what the textbooks don't tell you. Bull. Econ. Res. 52 (3), 0307–3378.
Tepe, M., Vanhuysse, P., 2013. Parties, unions, and activation strategies: the context-dependent politics of active labor market policy spending. Polit. Stud. 61 (3), 480–504.
Tepe, M., Vanhuysse, P., 2014. Taking social policy personally: the effect of personality traits and regime socialization on welfare state attitudes in Germany, R&R, European Political Science Review.
Thoenig, M., Verdier, T., 2003. A theory of defensive skill-biased innovation and globalization. Am. Econ. Rev. 93 (3), 709–728.
Tinbergen, J., 1975. Income Distribution: Analysis and Policies. American Elsevier (New York), North-Holland (Amsterdam).
Tóth, I.Gy., 2014. Revisiting grand narratives of growing income inequalities: lessons from 30 country studies. In: Nolan, B., Salverda, W., Checchi, D., Marx, I., Mcknight, A., Tóth, I.Gy., van de Werfhorst, H. G. (Eds.), Changing Inequalities and Societal Impacts in Rich Countries: Thirty Countries' Experiences. Oxford University Press, Oxford, pp. 11–47.
Tóth, I.Gy., Keller, T., 2013. Income distribution, inequality perception and redistributive preferences in European countries. In: Gornick, J.C., Jäntti, M. (Eds.), Economic Inequality in Cross-National Perspective. Stanford University Press, Stanford, pp. 173–220.
Tóth, I. Gy, Medgyesi, M., 2011. Income distribution in new (and old) EU member states in Corvinus. J. Sociol. Soc. Policy 2 (1), 3–31.
Tóth, I.Gy., Horn, D., Medgyesi, M., 2014. Rising inequalities: will electorates go for higher redistribution? In: Salverda, W., Nolan, B., Checchi, D., Marx, I., Mcknight, A., Tóth, I.Gy., van de Werfhorst, H.G. (Eds.), Changing Inequalities and Societal Impacts in Rich Countries: Analytical and Comparative Perspectives. Oxford University Press, Oxford, pp. 195–217 (Chapter 8).
Tsai, P.L., Huangy, C.H., Yang, C.Y., 2012. Impact of globalization on income distribution inequality in 60 countries: comments. Global Econ. J. 12 (3), 1–10. Article 7, http://www.bepress.com/gej/vol12/iss3/7.
Van Kerm, P., 2007. Extreme Incomes and the Estimation of Poverty and Inequality Indicators from EUSILC. IRISS Working Paper Series, CEPS-Instead, Luxembourg, p. 51.
Van Kerm, P., Pi Alperin, M.N., 2013. 2013: inequality, growth and mobility: the intertemporal distribution of income in European countries 2003–2007. Econ. Model. 35, 931–939.
Van Reenen, J., 2011. Wage Inequality, Technology and Trade: 21st Century Evidence. CEP Occasional Paper No. 28, London.
Vanhuysse, P., 2002. Efficiency in politics: competing economic approaches. Polit. Stud. 50 (1), 136–149.
Vecernik, J., 2010. Earnings disparities and income inequality in CEE countries: an analysis of development and relationships. Luxemburg Working Paper, Luxemburg. Retrieved from, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1806991.
Verhoogen, E., 2008. Trade, quality upgrading and wage inequality in the Mexican manufacturing sector" Q. J. Econ. 123 (2), 489–530.
Ward, T., Lelkes, O., Sutherland, H., Tóth, I.Gy. (Eds.), 2009. European Inequalities: Social Inclusion and Income Distribution in the European Union, Tárki, Budapest, 214 pp.
Weeks, J., 2005. Inequality Trends in Some Developed OECD Countries, DESA Working Paper No. 6, UN.
Wheeler, C.H., 2005. Evidence on wage inequality, worker education, and technology. Fed. Reserve Bank St. Louis Rev. 87 (3), 375–393.

Whiteford, P., 2008. How much redistribution do governments achieve? The role of cash transfers and household taxes. In: OECD, Growing Unequal? OECD Publishing, Paris, pp. 97–121 (Chapter 4).
Wood, A., 1994. North-South Trade, Employment and Inequality: Changing Fortunes in a Skill-Driven World. Clarendon Press, Oxford.
Wood, A., 1998. Globalisation and the rise in labour market inequalities. Econ. J. 108 (450), 1463–1482.
Zhou, L., Biswasy, B., Bowlesz, B., Saunders, P., 2011. Impact of globalization on income distribution inequality in 60 countries. Global Econ. J. 11 (1), 1–16, Article 1.

CHAPTER 20

Globalization and Inequality

Ravi Kanbur
Cornell University, Ithaca, NY, USA

Contents

Abstract

To what extent can increasing inequality be explained by globalization? And if there is a connection, what if anything can and should be done about it? This chapter begins with an overview of how conventional trade theory has fared in predicting changes in inequality and how it has needed to be extended and expanded when, contrary to some received wisdom, greater global integration is associated with increasing inequality in developed and developing countries. From there, the chapter goes well beyond these concerns to take in the effects of crises on inequality, globalization and gender inequality, openness and spatial inequality, and the effect of international migration on inequality. Finally, reviews of the latest developments in the design of national and global policy to address the challenges of globalization and inequality are presented. The literature reviewed is lively and flourishing. Having animated the economic analysis and policy discourse for the past half century, the globalization–inequality nexus seems set to continue in this vein in the coming decades.

Keywords

Globalization, Inequality, New trade theory and inequality, Crises and inequality, Globalization and gender, Spatial inequality, International migration and inequality, National policy, Global policy

JEL Classification Codes

D31, D33, F15, F24, F61, F63, F68, 019, 024

Handbook of Income Distribution, Volume 2B
ISSN 1574-0056, http://dx.doi.org/10.1016/B978-0-444-59429-7.00021-2

20.1. INTRODUCTION

Globalization is the dominant economic phenomenon of the last 30 years. Openness in trade, investment, and financial flows has grown dramatically. Inequalities within countries have also increased significantly during this period.[1] The natural question to ask is whether there is a connection between the two. To what extent can the increase in inequality be explained by globalization? And, if there is a connection, what if anything can and should be done about it?

Any exploration of inequality must begin by specifying inequality of what and inequality between whom. The focus in this chapter is on income inequality, although quite often measurement will be confined to inequality of consumption expenditure. As for inequality between whom, this can be between all individuals in the world, between nations, between individuals within nations, or between broad groupings within the nation. The focus of this chapter is inequality within developing nations. However, this in no way suggests that globalization is unimportant to inequality in developed countries. Evidence on inequality in developed countries is also referred to as relevant throughout this chapter. The inequality considered is primarily between individuals, but inequality between broadly defined groups within the nation (spatial and gender)—will also be discussed. The measure of inequality, which determines what aspect of the income distribution is emphasized, is also a relevant consideration. For the most part, this chapter considers standard measures of inequality such as the Gini coefficient.

A simple framework for linking income distribution and globalization is to write income as derived from different assets and the return on those assets plus net transfers. The transfers can be further disaggregated into private and public transfers. Assets can be disaggregated into basic factors such as land, labor, and capital, although further disaggregation, especially of labor between different skill levels, is also sometimes useful. The assets of an individual are therefore the capital and land that individual owns, plus the human capital embodied in that individual's labor power. The evolution of income distribution can then be decomposed into the evolution of assets, the evolution of rates of return to these assets, and the evolution of public and private transfers.

As noted above, different economic dimensions of globalization can be measured by increases in trade, investment, financial flows, and migration across national borders. These are of course outcome variables, determined by more fundamental causal variables such as natural endowment differences between nations and national and global policy. The literature often slips into the practice of labeling increases in trade, for example, as the causal factor whose consequences for inequality need to be investigated. This chapter is not immune to this tendency, but the caveat must always be borne in mind.

The focus of the large and growing literature that looks to uncover the links between globalization and inequality is primarily through the effects of globalization on rates of

[1] Chapter 9 of this volume covers trends in income inequality in developing and emerging economies, while Chapter 8 is devoted to inequality trends in developed countries.

return to assets, holding fixed the asset distributions. Even when assets are considered as mobile, the focus is on the impact of this mobility on returns to assets rather than on the distribution of assets. Within the analysis of returns, the literature is structured around gaps in returns to capital as a whole and labor as a whole, and around the gaps in returns to skilled and unskilled labor. The underlying assumption is that a widening in these gaps will increase interpersonal inequality as measured through standard indices such as the Gini coefficient. Because individuals who get their income primarily from capital generally have higher incomes than those who get their income primarily from labor and because skilled individuals generally have higher incomes than unskilled individuals, this is not an unreasonable assumption to make. However, it should be stressed that one cannot read directly from factor returns to the inequality of personal incomes. The distribution depends not only on factor prices but also on quantities. Nevertheless, in much of the literature, an analysis of inequality is replaced by an analysis of differentials in returns to capital and to labor at different skill levels.

Once the market distribution of income is determined, public and private transfers will contribute to the outcome of the final income distribution. These can be equally important as determinants of inequality, and globalization can affect them as well. First, international remittances, a natural consequence of international migration, can affect inequality in developing countries. Second, the greater ability of capital and high-income labor to cross borders can also have an impact on the progressivity of public tax and transfer regimes and thus on final inequality. This channel from globalization to inequality also needs to be considered.

With this background, the structure and plan of this chapter is as follows: Section 20.2 begins with the state of play in the three decades after World War II, from the 1950s through to the 1970s. The focus here will be on how the distributional predictions of the Heckshe–Ohlin (H–O) model, particularly the Stolper–Samuelson theorem, meshed with the great policy debates of the time, especially around the significance of the East Asian experience. These economies delivered a "growth with equity" miracle in a regime of trade openness at a time when other economies with import substitution regimes were either stagnating with low growth rates (like India) or were growing but with high and rising levels of inequality (like Brazil). This experience was consistent with the prediction that in economies that were abundant in unskilled labor, opening up would lead to a narrowing of the gap between unskilled labor on the one hand, and skilled labor and capital on the other hand. The East Asian experience was crucial to informing the debate and to persuading the international financial institutions and, in turn, many developing country governments to open out their economies in the 1980s and 1990s.

Section 20.3 provides a thumbnail sketch of the evolution of within-country inequality in the 1980s, 1990s, and 2000s, with a particular focus on the impact of openness.[2]

[2] This brief review complements the more detailed discussion of post-1970 trends in inequality in developing countries in Chapter 9.

The bottom line is that openness seems to have been associated with increases in pretransfer inequality. Clearly, this pattern from the 1980s onward questions the validity of the basic H–O framework in explaining the inequality consequences of trade, especially because inequality rose both in economies that were relatively labor abundant and in those that were relatively labor scarce. The section then turns to a range of new theories, particularly those emphasizing heterogeneity of workers and firms and market-based selection effects intensified by trade. Such a perspective, it turns out, is more successful in explaining the stylized facts of openness and inequality in the last three decades.

Sections 20.2 and 20.3 focus on a particular notion of openness (greater levels of trade and cross-border investment), a particular entry point to income distribution (differential rates of return to broadly defined factors of production), and a particular notion of inequality (between persons within nations). These are of course major strands in the literature. However, the remaining sections of the chapter take up a number of extensions, modifications, and generalizations that have developed in the last few years from this base.

Section 20.4 focuses on an aspect of globalization that became prominent with the East Asian crisis in 1997 and occupied policy makers' thinking strongly in the 2008 global financial crisis. How do crises induced by globalization of financial flows affect inequality within countries? There is significant literature developed on this topic based on country studies and global analysis for the crises of the 1990s and the 2000s. This section will review this literature and take stock.

Section 20.5 takes up a particular dimension of inequality—gender inequality.[3] This is an important aspect of inequality in its own right, with substantial and significant literature focusing specifically on globalization and gender inequality. For example, the Bangladesh garment sector or the Mexican maquiladoras employ women disproportionately, and there is heated debate about the conditions of work in these sectors and whether the women are better off here compared to the best alternative.[4] The empirical literature matches the policy debate, supporting both sides of the argument, and will bear a systematic review to draw out the main analytical issues and "centre of gravity" of the conclusions.

Section 20.6 addresses a dimension of inequality that is prominent in the policy discourse—spatial inequality within a country. This can be seen merely as a component or a contributor to interpersonal inequality, but doing so would miss important recent analytical and policy strands in the literature—for example, how agglomeration economies interact with openness, or the political economy of uneven development within a country.

Section 20.7 begins the assessment of openness, transfers, and inequality by looking at private transfers through remittances. It also takes up the more general question of the

[3] The general question of gender and inequality is taken up in Chapter 12.

[4] This also highlights the difference between income and broader measures of well-being, including for example health and safety standards.

impact of international migration on inequality in developing countries. Can migration and remittances exacerbate domestic inequality? There is some evidence that it can, and this may be a contributory factor in the association between global integration and within-country inequality.

Section 20.8 moves to public transfers and public policy in general and asks how greater mobility of capital and skilled labor in particular may constrain governments from pursuing progressive tax and transfer policies with consequences for inequality in the final distribution of income. This section also takes up the more general question of international coordination of public policy to address the impact of openness on inequality.

Section 20.9 concludes the discussion with suggestions for areas of further research. A final caveat is in order, however. This chapter is about globalization and inequality, and the focus is naturally on the links from globalization to inequality. As such, it may sometimes give the impression that globalization is the main factor behind inequality increase. There are of course other forces affecting inequality, and trade and capital flows may not even be the most important factors, although they surely interact with and influence a range of structural and policy influences on inequality.

20.2. IMMEDIATE POST-WAR THEORIES, PREDICTIONS, AND EVIDENCE

Although economic historians have been interested in the links between globalization and inequality in the nineteenth and early twentieth centuries,[5] we begin this discussion by considering the first three decades after World War II. At the start of this period, much of the development literature was focused away from global opportunities. It was either concerned primarily with domestic processes to the neglect of the global context, or it was suspicious of international trade, investment, and capital flows.

An example of a theory of development that was isolated from global forces is the classic Lewis (1954) surplus labor perspective. In the first part of this paper, Lewis analyzes a pure closed economy in terms of drawing labor away from the traditional surplus labor sector toward modern capitalist forms of production, a process that continues until labor becomes scarce and wages start rising.

What is interesting and not very well appreciated, however, is that the Lewis (1954) paper was in two parts. Part II of the paper deals with the open economy in the phase when surplus labor is exhausted:

> *When capital accumulation catches up with the labour supply, wages begin to rise above the subsistence level, and the capitalist surplus is adversely affected. However, if there is still surplus labour in other countries, the capitalists can avoid this in one of two ways, by encouraging*

[5] See for example Lindert and Willamson (2001).

immigration or by exporting their capital to countries where there is still abundant labour at a subsistence wage.

Lewis (1954, p. 176)

Lewis carries out a detailed analysis of a number of archetypical cases of trade and investment. Among his conclusions are the following:

The export of capital reduces capital formation at home, and so keeps wages down. This is offset if the capital export cheapens the things which workers import, or raises wage costs in competing countries. But it is aggravated if the capital export raises the cost of imports or reduces costs in competing countries.... The importation of foreign capital does not raise real wages in countries which have surplus labour, unless the capital results in increased productivity in the commodities which they produce for their own consumption.... The Law of Comparative Costs is just as valid in countries with surplus labour as it is in others. But whereas in the latter it is a valid foundation of arguments for free trade, in the former it is an equally valid foundation of arguments for protection.

Lewis (1954, p. 189)

This perspective on openness dovetailed with other perspectives such as export pessimism on the demand for products produced by developing countries. Many models of development at this time were built on this foundation. Overall, it would be fair to say that Lewis was indeed suspicious of openness in trade and investment raising wages relative to the return to capital in a country with surplus labor. In addition, in his other writings, he was quite "Kuznetsian" in seeing the initial stages of development as leading to rising inequality because, as he said (Lewis, 1976),

Development must be inegalitarian because it does not start in every part of the economy at the same time.... There may be one such enclave in an economy, or several; but at the start development enclaves include only a small minority of the population. (p. 26).

Thus, as opportunities opened up for trade they would be taken by some and not others and this would create inequality. At the same time, surplus labor would prevent the narrowing of the inequality on average between labor and capital. Overall, then, a pessimistic view exists on globalization and inequality.

Counter to this perspective is a view of the world without surplus labor, with trade between economies with different degrees of labor scarcity. This neoclassical H–O model famously leads to the "Stolper–Samuelson" conclusion that opening up of trade will raise the relative return of the relatively abundant factor. Because in developing countries this factor is labor relative to capital, it must follow that opening up will narrow the differential in rates of return to labor and capital. Making the reasonable assumption that owners of capital are richer than those who earn their living through their labor power, it follows that globalization will reduce inequality in developing countries.

These theoretical perspectives corresponded, of course, to policy stances. Most developing countries in the immediate post-war period adopted import substitution strategies—convinced that opening up would be bad for growth and for inequality.

Elaborate multisector planning models, such as those for the first Indian 5-year plans, had these key elements of a focus on domestic markets and domestic industrialization. Latin American countries adopted import substitution strategies, as did the newly independent African countries in the 1960s and 1970s. However, a group of countries in East Asia went against this trend and, from the 1960s onward, pursued policies of integration with the global economy. There is of course a huge debate on the details of these strategies. In particular, there is debate on the extent to which their policies can be classified as "free market" policies. However, there is no question that for three decades after the war, these economies, in contrast to other economies discussed above, did indeed integrate into the world economy in a purposive manner.

The East Asia experience was crucial to the policy debates of the 1970s and 1980s and to the turn in policies that one began to see in the rest of the developing word from the 1980s and 1990s onward. The 1960s and 1970s saw what has been dubbed the "East Asia miracle" of growth with equity. Not only did this group of countries have historically high growth rates, and higher growth rates than their contemporaries, they also managed growth with falling levels of inequality. The combination of high growth and falling inequality meant a sterling record in poverty reduction as well.

Of course the details of the inequality performance are varied and are not quite so uniform across countries and over time. There were periods of increasing inequality in some of the countries, and there were differences between Northeast Asia and Southeast Asia.[6] But there is a general acceptance that the East Asia story is one of growth with inequality kept in check. However, the interpretation of the facts is a different story. Already alluded to is the use of experience to support both the "free market" and the "judicial intervention" strands of the policy debate. The distributional outcomes have similarly been interpreted in different ways. One straightforward interpretation is in terms of support for the neoclassical H–O model with its prediction that opening up would narrow the returns to labor and capital and, with it, bring about a reduction in inequality. Indeed, this was the interpretation that was most used by those urging other countries, like India, to adopt outward-oriented policies. Thus, the classic exposition by Bhagwati and Desai (1970) represents a turning away from the nostrums of the immediate post-war, post-independence consensus in India that equitable development could only be achieved through import substitution and industrial planning. This strand of literature found its apogee in a series of studies by the World Bank in the 1980s, including for example, in Papageorgiou et al. (1990), the capstone to publications entitled the "Liberalizing Foreign Trade Series." The contrast of East Asia with stagnation in India and growth with inequality in Brazil was very much highlighted in this literature. At the same time, the integration of Europe through the European Union, and the success it

[6] World Bank (1993) and Jomo (2006).

delivered over a long period of high growth with falling inequality in the immediate post-war decades, was also relevant in the policy discourse.

However, the East Asian experience has also been used to support the thesis that the equity dimensions of outcomes owe a significant amount to other structural and policy features. Among these are the land reforms instituted by the occupying American forces in South Korea in the 1940s and 1950s, which meant that they entered the next phase of development, in the 1960s and 1970s, with supportive initial conditions for equitable development. Further, in these countries and in other East Asian countries, proactive policy had ensured a very wide spread of basic education. Here is how Adelman (n.d.), the leading scholar of South Korean development strategy at that time, sets out these structural factors in the country from the end of World War II until the beginning of the 1960s:

> *There were two waves of land reform, in 1947 and 1949. In 1947, the U.S. military government decreed that the land confiscated from Japanese farmers and Japanese corporations should be redistributed to tenants.... The second wave of land reform redistributed the holdings of Korean landlords owning more than 3 chongbo (7.5 acres or about 3 hectares) to tenant farmers and landless farm laborers.... The distribution of land holdings became very even.... The bulk of government investment during this period was on social development...Over this period, the literacy rate increased from 30 to over 80 percent.*

These structural factors have to be seen in conjunction with the perspective of Lewis (1976) that initial differences in advantage can be magnified by the appearance of economic opportunity. Thus, perhaps the best interpretation of the East Asia experience is being supportive of both a structuralist view and a neoclassical perspective based on the H–O model. The land reforms and the wide spread of education simultaneously reduced surplus labor while at the same time making the distribution of assets (land and human capital) much more equal. The stage was thus set for an opening up and integration into the global economy to deliver growth with equity. However, the outcome was dependent on the initial conditions at the time of the opening up, conditions that need not necessarily hold in other countries, or at other time periods.

20.3. EXPERIENCE AND NEW THEORY FROM THE 1980s ONWARD

The policy debates of the three decades following World War II influenced and were influenced by the analytical frameworks developed to understand the impact of trade and investment openness on inequality. The experiences of this period, in particular the perceived "growth with equity miracle" of East Asian economies, contrasted with the stagnant or rising inequality in countries such as India (with relatively low growth) or Brazil (with relatively high growth), were particularly important in convincing policy makers to open up their economies from the 1980s onward. However, the importance of structural features such as the low degree of asset inequality in East Asian economies when they launched their drive to openness seems not to have received as much attention. The past three decades have been periods of ever intensifying globalization as measured by trade integration and the magnitude of capital flows. What has been the experience with inequality?

The experience of inequality in the United States (and other developed economies) is interesting because of the possible light it can shed on the predictions of the standard H–O model. The simple model has the powerful prediction that opening up will narrow the returns between labor and capital in countries with a relatively low capital-to-labor ratio, or between skilled and unskilled labor in countries with a relatively low skilled-to-unskilled labor ratio. The observance of these trends in East Asia was read as support for the model. The flip side of this same prediction is that the gap between these returns should widen in countries with relatively high ratios of capital to labor and of skilled labor to unskilled labor. This did not happen in the United States in the 1960s and 1970s, but it has been happening since the 1980s. Now, it can be argued that given the relative size of the U.S. economy, it was only in the 1980s and 1990s, with the opening up of China and India, that the trade effects could be felt strongly enough to impact factor returns. So, the inequality trends in the United States could indeed be claimed as partial support for the H–O model.

There is, however, the issue of how much of the rising inequality in the United States can be attributed to trade, and how much to other factors, specifically to technology. The overview by Pavcnik (2011) captures the recent consensus:

> *A large body of research on this topic finds little support that international trade in final goods driven by relative factor endowment differences can account for much of the observed increase in skill premiums in developed and developing countries.... First, the Stolper–Samuelson mechanism suggests that increased relative demand for skilled labour in countries abundant in skilled labour occurs as a result of shifts in the relative demand for skilled labour across industries.... However, the employment shifts across industries have not been sufficiently large to account for the large increase in wage inequality. Most of the observed increase in demand for educated labour in countries such as the United States is driven by increased relative demand for skilled labour within industries. (p. 242)*

There is significant debate on the relative role of trade. Although Krugman (2008) argues against his own earlier view that trade was a relatively small factor in explaining the rise of inequality compared to technology, there are also criticisms of the "small role of trade" view by Irwin (2008), Katz (2008), and Autor (2010). It would be fair to say that skill-biased technical change is considered to be a major driving force, if not necessarily the dominant force, behind rising inequality.[7] This empirical and policy debate has in turn

[7] Of course, this argument on skill-bias depends on the magnitude of the elasticity of substitution between skilled and unskilled labor; Further, as Atkinson (2008) points out, with a supply response a rise in the rate of skill-biased technical progress leads only to a higher level of inequality, not permanently rising inequality.

fed into an emerging literature that goes beyond simple H–O/Stolper–Samuelson formulations to consider within-industry wage differentials between heterogeneous firms and how these could be affected by trade.

The H–O predictions on trade and inequality could be argued to have been confirmed by the experience of rising trade and falling inequality in East Asia in the 1960s and 1970s. They could equally be argued to have been confirmed by the experience of the rising trade and rising inequality in the United States from the 1990s onward, although there is consensus that the forces of technology provide stronger explanation. However, the difficulty for the H–O model is that, contrary to its prediction, and contrary to the experience of East Asia in the 1960s and 1970s, from the 1980s onward, the experience of Asian economies and that of Latin America until the 2000s has been one of rising trade and rising inequality. As the comprehensive review by Goldberg and Pavcnik (2007) concludes:

> *The survey of the evidence confirms Wood (1999), who noted that inequality increased in several middle-income Latin American countries that liberalized their trade regimes in the 1980s and 1990s. It further suggests that this positive relationship holds in the cases of India, China and Hong Kong. As noted previously by Wood (1999), the experience of developing countries that globalized during the 1980s and 1990s contrast with the experiences of several Southeast Asian countries (South Korea, Taiwan, Singapore) that underwent trade reforms in the 1960s and 1970s. The latter underwent a decline in inequality as they opened up their economies to foreign markets. (p. 54)*

A number of comments are in order before we proceed to discuss the implications of these facts for the H–O model. First, although the economies of Latin America liberalized during the 1980s and 1990s, this was also a period of painful macroeconomic adjustments and slow downs, and this could confound attribution of the causes of inequality. Second, note that the simple Lewis–Kuznets model discussed in the last section could indeed still predict an increase in inequality with opening up. Finally, two major further stylized facts have been established since the Goldberg and Pavcnik (2007) survey. First, inequality also increased in East Asia in the 1990s and the 2000s.[8] Second, inequality has declined in Latin America since the 2000s.[9] Both of these are after their major periods of trade liberalization and, particularly in Latin America, have been linked to redistributive policy—these policy issues will be taken up in a subsequent section.

The basic H–O/Stolper–Samuelson framework is foundational in the discourse on trade and inequality. However, questions about its validity have been raised by the finding that inequality in many developing countries has increased since the 1980s despite increases in trade. This disconnect between prediction and outcome has led to a fruitful search for alternative explanations of why an increase in trade may increase inequality, and some of the theories advanced have also been helpful in understanding the impact

[8] Kanbur and Zhuang (2012).
[9] Lustig et al. (2011).

of trade on inequality in developed countries as well. In this section, we examine a range of such theories as illustration of the direction the literature is taking in light of the experiences of the last three decades.

In the wake of the failure of the basic two goods, two factors H–O model to predict co-movement of trade and inequality, a range of models were developed that vary the technology or number of factors and goods (including the introduction of nontraded goods) in order to derive predictions more consistent with the data. Thus, for example, Wood (1994) moves from the two-factor model with a skilled/unskilled labor division to consider a three-factor model with workers classified as skilled (high education)/semi-skilled (basic education)/unskilled (no education). Further, there are three types of production—skill-intensive manufacturing, semi-skilled intensive manufacturing, and agriculture. In this setting, for a country with comparative advantage in agriculture we get the standard prediction that opening up will reduce inequality. However, for countries with a relatively large number of semi-skilled workers, opening up will increase their wages relative to the wages of both high-skill and unskilled workers. The effect on inequality is thus ambiguous, and measured inequality could increase. While an interesting extension to the basic H–O model, it is not clear how well this fits the data. After all, East Asia in the 1960s could be argued to be a region with a predominance of basic education, and evidence from the 1980s onward suggests that wages of the highly skilled have risen disproportionately.

In the same spirit, Davis (1996) considers a two-factor (he calls them capital and labor), three-goods H–O model, with market imperfections that prevent factor price equalization and full diversification of production. The three goods differ in the capital intensity of production technology. With countries ranked by capital intensity of factor endowment, the least developed countries will export the least capital intensive commodity and import the next most capital intensive. For these countries, the standard result will hold—opening up will narrow the gap in factor returns. However, for countries with intermediate levels of capital intensity of factor endowment, which will export the commodity with intermediate capital intensity of production and import the commodity with highest capital intensity of production, opening up will have the opposite effect. Of course, for the most developed countries we again have the standard Stolper–Samuelson result. At least for developing economies at intermediate levels of capital intensity, then, this type of theorizing might explain co-movement of trade and inequality. Such countries might, in principle, include East Asia from the 1980s onward and Latin America at the time of its opening up in the 1980s and 1990s.

The papers by Wood (1994) and Davis (1996) are examples of attempts to predict co-movements of trade and inequality within a recognizable H–O framework but with more disaggregated specification of commodities or factors. This trend has continued in the literature, with added complications such as capital-skill complementarity in production—to the point that the discourse of today cannot really be labeled as a

H–O discourse. In what follows, I will consider the literature that highlights heterogeneity of workers, firms, and production processes.

Helpman et al. (2010) bring together several strands of the modern trade literature with a focus on firm and worker heterogeneity and derive predictions on trade and inequality that are consistent with many of the empirical findings of the last 30 years. Following Melitz (2003), the model supposes heterogeneous firms producing differentiated commodities. Firms can enter by paying a fixed cost, but discover their productivity only after paying the sunk cost. The productivities are drawn from a Pareto distribution, an assumption that helps the tractability of the model. After productivity is revealed, firms decide whether and how much to produce for export, for the domestic market, or for both, or they exit altogether. Production involves a fixed cost, and output is a function of firm productivity, number of workers hired, and their average ability. A specific functional form is used for tractability, but the key aspect is that these three elements are complementary to one another.

Worker ability is also assumed to have a Pareto distribution—again for tractability. Search and matching frictions exist in the model, and firms can pay more to match with more workers. Further, among the workers the firm can screen for higher abilities above a cutoff by paying a cost (with a higher cutoff costing more), but it cannot distinguish abilities beyond this cutoff. Thus, all workers in a firm are paid the same wage. The wage is modeled as emerging from the outcome of a bargaining game between the firm and the average worker.

Fixed costs of production, and fixed costs of exporting, mean that firms with very low productivities do not produce at all, while firms with high productivities select exporting because of the existence of a cost of trading. Given costs of search and screening, it can also be shown that firms with higher productivity and revenue search more and use a higher ability cutoff, so that they have higher ability workers on average and thus higher wages. The key point is that exporting firms pay higher wages in equilibrium. Thus, if we start from autarky, where fixed costs of exporting are so high that nobody exports, and reduce these fixed costs in a comparative static manner so that some firms begin to export, wage inequality is introduced where none existed before. This applies to all countries; thus, opening up can increase inequality in all countries—developed and developing—because of the selection effects of exporting.

Verhoogen's (2008) is another example of a similar model where selection effects can explain co-movement of trade and inequality. The idea here is that exporting requires the production of higher quality products and only the most productive will find it profitable to go into exporting. With a mechanism of higher wages in more productive firms, this leads to greater inequality with more openness. It should be noted that the Helpman et al. (2010) model also has an intriguing result at the other end of the spectrum where exporting costs are so low that all firms export. Then, once again, wages are equal. In their model, inequality first increases and then decreases as opening up intensifies—an "inverted-U" relationship between inequality and openness. It is of course an empirical question as to whether the intensified globalization from the 1980s onward has now taken some countries to the point where the model would predict falling inequality. If this was the case for some countries, of course, the model could not explain the co-movement of trade and inequality for those countries, and other explanations would have to be considered.

A selection mechanism of a different sort is present in studies of outsourcing as exemplified by Feenstra and Hanson (1996, 1997), which also relates to a broader literature in outsourcing and FDI in trade. They considered a scenario where the final output is produced using intermediate inputs that are produced using different intensities of skilled and unskilled labor. Consider now two economies with different endowments of skilled and unskilled labor. For any given pattern of trade costs, the skilled labor abundant (developed) economy will use the more skilled intensive production of intermediate inputs. When trade costs are lowered in a comparative static exercise, some of this production is relocated from the developed economy to the developing economy. However, the activity that is relocated is the least-skilled intensive in the developed economy and the most-skilled intensive in the developing economy. This increases skill intensity of production in both the developing and the developed economy and, hence, widens the wage gap between skilled and unskilled labor in both economies. Feenstra and Hanson (1997) show empirical support for this as explaining rising wage inequality in Mexico.

Feenstra and Hanson (1997) highlight an aspect of globalization that has come to the forefront in the last 30 years, namely, foreign direct investment (FDI). The issue of portfolio and financial flows will be discussed in a subsequent section, but longer term FDI has also been important in the recent growth surges in developing countries. What are the implications of FDI for inequality?

The theory of FDI in the simple Lewis model discussed in the previous section suggests that as wages rise in a former surplus-labor economy, capitalists will look to investment opportunities abroad, presumably in economies where wages are lower still. If these economies are themselves in a state of surplus labor then further investment will raise the share of capital and worsen the distribution of income for that reason. However, if the "Lewis turning point" has already been reached in the economy receiving FDI, this investment will raise wages further in that economy, and this could be a channel for reducing inequality.

Modern theories of the impact of FDI build on the H–O framework and then bring in firm and worker heterogeneity, as in the analysis of Feenstra and Hanson (1997). Overall, it would be fair to say that the theoretical conclusions are ambiguous, with some suggestion of FDI contributing to an increase in inequality in developing countries at the start of the process, with a possible turnaround in the later stages. For example, Figini and Gorg (1999) discuss the transition as domestic firms absorb the new technology of the FDI.

Inequalities may be created in the early stages, but are mitigated in later stages as the transition proceeds—an inverted-U relationship has framed much of the empirical work in this area. The large and growing empirical literature also gives mixed results, with perhaps a greater weight to the conclusion that FDI is associated with rising inequality in earlier stages,[10] but that there may be a turnaround, and that the impact is muted or even negative at higher levels of income per capita.[11]

Selection effects as the result of global integration are now central to the trade and FDI literature and, thus, to the attempts to explain co-movement of trade and inequality. They do appear to provide a coherent explanation of increases in inequality in both developed and developing countries, and for this reason, they merit close theoretical and empirical attention in the years to come.[12]

20.4. ECONOMIC CRISIS AND INCOME DISTRIBUTION

It is often said that globalization brings risks as well as opportunities at the macroeconomic level. Greater integration with the global economy can lead to the economy being buffeted by global fluctuations in trade and capital flows. What has been the contribution of openness to macroeconomic volatility? The current consensus and weight of research seems to suggest that openness is associated with greater volatility (Bekaert et al., 2006; Easterly et al., 2001; Kose et al., 2006; Rodrik, 1997).[13] The paper by Di Giovanni and Levchenko (2008) conducts a careful analysis of the channels through which trade openness increases volatility. They test for three channels: (i) increased volatility of individual sectors, (ii) increased co-movement of sectors, and (iii) a more specialized production pattern. They find support for the first and third but find that more openness in a sector

[10] Feenstra and Hanson (1997) for Mexico, Figini and Gorg (1999) for Ireland, Taylor and Driffield (2005) for the United Kingdom, Tsai (1995) for a cross-section of 33 countries and Basu and Guariglia (2007) for 8 countries, find the increasing relationship.

[11] Figini and Gorg (2011) find for a cross-section of 100 developing and developed countries that: "Results for developing countries are robust and suggest the presence of a nonlinear effect: wage inequality increases with FDI inward stock, but this effect diminishes with further increases in FDI. For developed countries, wage inequality decreases with FDI inward stock, and there is no robust evidence to show that this effect is nonlinear." (p. 1473)

[12] The literature on heterogeneous workers, heterogeneous firms, and trade is exploding and it would be impossible to do it justice in the space available. The recent surveys by Grossman (2013) and Costinot (2009) are useful. The paper by Costinot essentially generalizes H–O to trade models with heterogeneous workers and firms.

[13] It should be noted that it was concern about exposure to foreign trade fluctuations that was very much behind policy debates about unemployment in the United Kingdom and elsewhere at the end of the nineteenth century, and this led to the introduction of social insurance.

reduces the co-movement of its growth with overall growth in the economy, which tends to reduce aggregate volatility. However, the overall effect of openness on volatility is clear:

> *. . .moving from the 25th to the 75th percentile in trade openness is associated with an increase in aggregate volatility of about 17.3% of the average aggregate variance observed in the data. The impact of openness on volatility varies a great deal depending on country characteristics, however. For instance, we estimate that an identical change in trade openness is accompanied by an increase in aggregate volatility that is five times higher in the average developing country compared to the average developed country. Lastly, we estimate how the impact of trade changes across decades. It turns out that all three channels, as well as the overall effect, increase in importance over time: the impact of the same trade opening on aggregate volatility in the 1990s is double what it was in the 1970s. (p. 5)*

However, a major focus of the last two decades has been volatility and crises induced by financial flows. Financial crises appear to be the new normal in the global economy. Fully fledged global crises, such as the one that occurred in 2008–2009, or the East Asian financial crisis of 1997, which also had global repercussions, are recognized as at least aided by the far greater ease of movement of portfolio capital around the world in the wake of capital account liberalizations from the 1990s onward. These global crises also have implications for national level macroeconomic volatility, which has also been affected by trade openness. Indeed, Hnatkovska and Loayza (2013) argue that the increased volatility can be attributed more to crises ("large recessions") than to the normal economic cycle.

There is now a consensus that volatility is associated with lower growth—Hnatkovska and Loayza (2013) present only the most recent assessment in this vein. However, this section will review the recent discourse on the consequences of economic crisis for the distribution of income—for poverty and for inequality.[14] The literature has set out a range of channels through which a global collapse of the type seen in 2008–2009, or the more limited contagion effects of the crisis in 1997, feeds through into income distribution. Atkinson and Morelli (2011) and Baldacci et al. (2002) highlight the following channels:

1. Economic slowdown. As a "balance sheet adjustment" recession takes hold in originating countries, it is transmitted through trade to other countries. Thus, each country faces an economic slowdown. There is unemployment in the formal sector and consequent downward pressure on earnings in the informal sector. We would expect the impact of economic slowdown to be rising poverty and also rising inequality.
2. Relative prices and sectoral effects. For a particular country, the decline in international demand may be concentrated in specific sectors, with quantity and price effects. Thus, unemployment and wage contraction will have sectoral patterns that differ

[14] There is a growing literature on whether inequality breeds crises—a good example of this line of argument is in Rajan (2011). We will not discuss this strand of the literature here.

from country to country. Here, the impact on wage inequality will depend on whether the sectors that are negatively impacted are the ones that were paying higher wages to begin with. If so, crisis could actually reduce inequality through this channel (although poverty would rise).

3. Asset effects. Changes in interest rates and revaluation of assets can affect incomes and wealth at the top of the income distribution. If there are major downward valuations and reductions in income from capital, then crises could reduce wealth and income inequality through this channel.
4. Policy responses. This includes the consequences of fiscal retrenchment, which will have impacts at the lower tail of the income distribution, or bank bailouts, which will affect the top end of the distribution. In general, fiscal retrenchment through reducing public employment, or support for public works schemes and other forms of unemployment support, would increase poverty and inequality in the social sectors. Bank bailouts would support asset values and incomes at the top end of the income distribution and increase inequality. Finally, an important channel linking crises and distribution is the drastic devaluation most often undertaken as a response to a balance of payment crisis. This is equivalent to a drop in real wages and an increase in profits.

Each of these channels can have multiple impacts on poverty and inequality, so the overall effect is an empirical question. Ravallion and Chen (2009) focus on the 2008 global financial crisis and provide projections of the likely impact on poverty. They estimate that "the crisis will add 64 million people to the population living under a dollar a day." The methodology for doing this, however, assumes no distributional change within a country, based on the observed regularity that "relative inequality falls about as often as it rises during aggregate economic contractions, with zero change on average." Thus, Ravallion and Chen (2009) simply apply projected contraction in total consumption and assume this contraction to be distributionally neutral. They do recognize, however, that "while distribution neutrality is plausible on average, there will be some countries where the poverty impact of the crisis is greater than these calculations suggest, and some where it will be smaller. Country-specific analysis would be needed to determine which countries might have above-average impacts."

An attempt at identifying the impacts of poverty and inequality through cross-country regression techniques is presented by Baldacci et al. (2002). They define crisis episodes, identify appropriate controls of country-time spells, and estimate the impact of crises on different dimensions of income distribution. Not surprisingly, they find that crises are associated with rising poverty. However, in terms of income distribution, they find that "The main losers in terms of changes in income shares are not the poorest (lowest income quintile) but those in the second (lowest) income quintile. The income share of the highest quintile also falls in crisis years relative to pre crisis years." Thus, treating this regression finding as a representation of the average outcome, the results are consistent with the assumptions of Chen and Ravallion who state that crises are, on average, distribution neutral.

The post-1997 crisis experience highlights the country-specific differences that can arise. Hagen (2007) argues that income inequality rose significantly in Korea after the crisis. Similarly, inequality rose in Singapore and Malaysia, but it fell in Indonesia and in Mauritius (Atkinson and Morelli, 2011). Atkinson and Morelli (2011) assess the association between crises and inequality for a large number of crises over a long period of time. They distinguish between banking crises and crises of collapse in consumption. They look at the time path of inequality on either side of the identified crisis. For the former, they conclude that "the empirical evidence suggests that cases in which inequality tend to increase following the crisis are in majority, although we should caution that the sample size is too limited to draw firm conclusions." For the latter, "empirical evidence concerning 'change in direction' suggests that consumption crises are more associated with reduction in inequality. No particular pattern stands out from the analysis of GDP crises."

It would seem, therefore, that no easy generalizations are available for the impact of crises on inequality, as might be expected from the multiple channels through which they can work and how initial conditions in a country can affect the impact. What this means is that we need country-specific modeling to analyze and to predict the impact of crisis on inequality. One such approach is that of a microsimulation model, as in the work of Habib et al. (2010). This approach combines macroeconomic projections with transmission mechanisms to the income distribution:

> *The model focuses on labor markets and migration as transmission mechanisms and allows for two types of shocks: shocks to labor income, modeled as employment shocks, earnings shocks or a combination of both; and shocks to non-labor income, modeled as a shock to remittances. Shocks can be positive or negative depending on the trends outlined by the macroeconomic projections. In most cases labor income and remittances account for at least 75-80% of household income. (p. 5)*

Such country-specific analysis can then be used both to identify early warning indicators and to design possible policy responses. For example, the authors apply the model to Bangladesh and recommend monitoring of remittances and wages by sector as indicators of the need for action.[15] A range of these models and methods is surveyed in Bourguignon and Bussolo (2012), and in Bourguignon et al. (2008). However, an important question arises as to whether we use anonymous distributions before and after crises or whether we use panel data, which follow individuals from before the crisis to after. Then, anonymous distributions can show no change even when there is considerable "churning" as a result of the crisis as pointed out by Robilliard et al. (2008).

[15] For an example of a microsimulation model to the impact of crisis on inequality for a developed country, see O'Donoghue et al. (2013).

20.5. GLOBALIZATION AND GENDER INEQUALITY

So far we have analyzed the relationship between globalization and interpersonal inequality without regard to the gender of the persons. Indeed, gender was not present in the classical developments in attempts to link trade theory to theory of income distribution. However, in the past quarter century this issue has come to the fore strongly in the policy and analytical literature. The analytical reasons for this development are related to greater evidence on gender dimensions of inequality and the development of nonunitary models of the household, which allows for the prospect of unequal outcomes within the household. The policy reasons are related to strong debates on whether the global integration of the past 25 years has hurt or helped women.

It is well established that there is a strong gender dimension to interpersonal inequality.[16] This is most easily demonstrated empirically for variables that can be quantified at the individual level. Patterns are country specific, of course. However, in many developing countries, educational attainments are lower for women than for men and especially so at lower incomes. Sex ratios at birth in some countries reveal discrimination against women in sex selection, and maternal mortality rates in many developing countries are at the levels that Sweden attained in 1900. Women earn less than men for similar work, but also women tend to work in sectors and occupations that are low paying.[17]

It is not easy to measure the magnitude of gender inequality along the standard dimension of consumption, because consumption data are usually collected at the household level in surveys. The first cut of measuring gender inequality by inequality between female-headed households and male-headed households is unsatisfactory for obvious reasons. The standard assumption in translating household level information into individual level well-being is to simply divide by household size and allocate per capita consumption of the household to each individual in the household. Of course, this suppresses all intrahousehold inequality including gender inequality. Thus, our standard measures of inequality are underestimates of true inequality because they set gender inequality in consumption within the household to zero. On rare occasions when individual level consumption data is available (for example on food consumption), it has been shown that the standard procedure understates inequality (and poverty) by as much as 25% (Haddad and Kanbur, 1990). Thus, gender inequality, as reflected in intrahousehold inequality, matters.

While it is accepted that gender structures inequality in an economy, there is less consensus on how exactly globalization interacts with this structure. How is the standard analysis of openness and inequality, for example, affected by structuring the economy along gender lines? And, overall, does globalization reduce gender inequality, or increase it?

[16] Chapter 12 of this volume is devoted to the topic of gender inequality.
[17] World Bank (2011), p. 74, 78, and 79.

Before looking at some evidence, let us consider how standard theoretical arguments on globalization and inequality could be modified by taking into account the gender dimension of production and income distribution. A standard piece of analysis in open economy macroeconomics is the effect of devaluation on the balance of payments. As is well known, the transmission mechanism is through "expenditure switching" brought about by raising the price of tradables relative to the price of nontradables. The distributional consequences of this have been analyzed in the usual way through the Stolper–Samuelson theorem. If tradables are relatively more intensive in their use of labor, then the relative return to labor will rise. Indeed, this was the argument made by many for the pro-poor and progressive aspects of devaluation.

However, suppose that tradables are actually more intensive in their use of male labor. Then, it is seen that male earnings will be favored. This should not matter much if there is perfect income sharing within the household—the representative household would gain overall if the policy of devaluation was efficient for the economy as a whole. However, if the household is not described by a unitary model, and if, for example, there is bargaining between the man and the woman and their outside options matter for the outcome of bargaining, then, the macro policy of devaluation will have the micro consequence of strengthening the bargaining power of males and will have a type of impact on inequality not contemplated in the classical analysis.[18] Of course, the outcome is context specific—it depends on which sector is male or female labor intensive. The main theoretical point, however, is that gender matters (Haddad and Kanbur, 1994).

The above is in terms of the pure demand for labor. However, there is also evidence that women are paid less for the same job. The impact of globalization on such wage differentials is uncertain. On the one hand, there is the standard argument that greater global competition will reduce the scope for discriminatory wage practices, and this should narrow wage differentials. However, to the extent that mobility of capital reduces bargaining power of workers, and to the extent that women are concentrated in industries where capital is more mobile, greater openness will lead to lower female wages (Seguino, 2007). The effects of this competition in footloose industries might be seen not just in standard wages but also in labor standards (Chau and Kanbur, 2003, 2006). Again, to the extent that women are disproportionately employed in such industries, the impact of globalization will affect them disproportionately.

There are two main empirical strands of the gender and globalization literature. The first is focused on the effects of openness on demand for female labor and on female wages. The second is related to the previous section—how crises affect women relative to men. We take up these strands one at a time.

[18] Again, there is considerable evidence that household decision making is not best described by the unitary model. For an early survey of the literature see Alderman et al. (1995). A recent review is provided in Chapter 16 of this volume.

The effects of opening up on the demand for female labor are nuanced and context specific. On the one hand, the demand for female labor rises through expansion of light manufactures. As the World Bank's World Development Report on Gender notes:

> *In the Republic of Korea, the share of women employed in manufacturing grew from 6 percent in 1970 to around 30 percent in the 1980s and early 1990s. . . . Similarly, in Mexico, female employment in manufacturing grew from 12 percent in 1960 to 17 percent in 2008, with 10 times more women in 2008 than in 1960.*
>
> ***World Bank (2011, p. 256)***

However, this phase contrasts with the next phase as there is a move to the production of more capital intensive goods (Seguino, 2013; Tejani and Milberg, 2010; Van Staveren et al., 2007). What about female-wage differentials? Here again, the evidence reflects the conflicting forces, which are resolved differently in different countries. As Seguino (2013) notes in her overview:

> *Evidence of the impact of trade and investment liberalization for gender wage equality is also mixed. Some studies show that gender wage differentials have declined, in large part due to narrowing educational gaps. But in several developing countries, including China and Vietnam, however, the discriminatory portion of gender wage gaps has increased. (p. 15)*

A final, newly emergent strand of the literature provides a gender perspective on the selection and heterogeneity models discussed in Section 20.3. The argument put forward by Juhn et al. (2013) builds on the idea that more productive firms enter into export and modernize technologies. If new technologies require less physical strength (the "brains" vs. "brawn" issue, as it is characterized in some circles), we would expect that demand for female labor would rise in blue-collar occupations and not in white-collar occupations. This is because new technology can change the "brain/brawn" mix in blue-collar occupations, but white-collar jobs will be unaffected on this score. The authors find that for Mexico, post-NAFTA tariff reductions are associated with rising female employment and wage shares in blue-collar jobs but not in white-collar jobs.[19]

The various contradictory forces are also highlighted in Bussolo and de Hoyos (2009). On the basis of their studies of Africa and Latin America, they conclude, essentially, that with forces pulling in opposite directions, the net effect of trade openness on gender inequality may well turn out to be fairly weak:

> *Overall, the messages of this volume are very clear: trade expansion exacerbates gender disparities in agricultural-based, African economies and reduces it in manufacturing-based economies like Honduras. . . . Admittedly, the magnitude of the links between trade shocks, producer prices, male versus female bargaining power, consumption decisions, future growth and poverty reduction does not seem too large. . . . To conclude, trade liberalization brings important gender effects, but the evidence collected here shows that these effects tend to be of a small and sometimes uncertain magnitude.*

[19] The brain–brawn issue is also discussed in World Bank (2011), p. 259.

While the literature on the trade effects of globalization on gender inequality thus renders a relatively neutral verdict, the same is not true of the literature on the impact of economic crises on women. The effects of economic downturns generally, and economic collapses in particular, are argued to be felt most sharply by women because they tend to be displaced first. In turn, they crowd into the informal sector, pushing down earnings further in that sector, which is in any case disproportionately female in employment (Braunstein and Heintz, 2008; Takhtamanova and Sierminska, 2009). It is further argued that the fiscal retrenchment that accompanies economic crises affects women disproportionately both directly and indirectly, through reducing public services that support women's work, such as health and child care (Seguino, 2013).

There is, finally, an intriguing and important, yet unresolved issue of the effects of globalization on societal norms that determine the structure of gender inequality. Based on the work of Kabeer (1997, 2000) and Hossain (2011), World Bank (2011) argues as follows:

> *In Bangladesh, the employment of hundreds of thousands of women in the ready-made garment industry feminized the urban public space, creating more gender-equitable norms for women's public mobility and access to public institutions. In the process, Bangladeshi women had to redefine and negotiate the terms of* purdah, *typically reinterpreting it as a state of mind in contrast to its customary expression as physical absence from the public space, modest clothing, and quiet demeanor.*

How widespread these effects are, and how much they can be attributed to globalization, is still under debate. What is clear is that any discussion of globalization and inequality must go beyond the classical analysis and develop theory and empirical investigation on globalization and the gender dimension of inequality.

20.6. OPENNESS AND SPATIAL INEQUALITY

The spatial dimension of inequality is a key concern in the policy discourse, because it intersects and interacts with disparities between subnational entities and jurisdictions. These entities sometimes have defined ethnic or linguistic characteristics, and in federal structures have constitutional identities that naturally lead to a subnational perspective on national inequality. This section considers the impact of globalization, in particular greater openness in trade, on spatial inequality.

What exactly is spatial inequality? One way of linking standard interpersonal measurement of inequality to regional inequality is to decompose national inequality into a between-region and a within-region component. The share of national inequality accounted for by the between-region component—which would be zero were it not for the fact that average incomes differ across regions—is then a measure of regional inequality. The fraction of total inequality accounted for by variation in average income across regions depends, of course, on the number of regions. The larger the number of regions, the greater the inequality that can be attributed to regional difference in mean income. Estimates vary, but 15–20% of spatial inequality in total inequality is not unusual (see Kanbur, 2006).

An alternative, however, is to consider the disparities in regional mean incomes directly, not weighted by their population. Equal weights correspond to some dimensions of many constitutions, where key elements of political power are divided equally between constituent provinces or states (Kanbur and Venables, 2005). In the case of just two entities, then, this could be simply the ratio of the two means, for example. For more than two entities, other standard measures of dispersion can be used. Yet, other measures are sometimes used in the literature, attempting to capture regional "polarization." However, as Zhang and Kanbur (2001) argue, such measures may not make that much difference in assessing trends.

Kanbur and Venables (2007) review the literature and provide other measures of the level of spatial disparities observed around the world. In particular, they highlight variations in poverty and human development indicators. In Africa, in 6 out of the 12 countries studied, the percentage of people below a poverty line constructed on the basis of information about households' asset holdings is more than 50% greater in rural areas than in urban areas. The smallest rural–urban difference is 30%. Similarly, school enrollments, and the ratio of girls to boys enrolled, is much higher in urban than in rural areas. In Peru, the incidence of poverty in districts at sea level was 46.1% in 1997, while for districts at an altitude greater than 3500 m above sea level it was 63.3%. In Indonesia in 1993, the rural poverty incidence was 46.5% in West Kalimantan, but only 10.7% in Yogyakarta.

However spatial inequality is measured, there are major differences in the literature on how much it should matter in policy design. One strand of the policy discourse can be characterized by the "balanced development" perspective, which holds that too much concentration of economic activity is inimical to equity and to efficiency. However, there is a contrary strand that is best expressed in the World Bank's World Development Report on Economic Geography (World Bank, 2008, p. 73):

> *For decades, "spatially balanced growth" has been a mantra of policy makers in many developing countries. It was an obsession of planners in the former Soviet Union.... And it has been the objective of governments of various political hues in the Arab Republic of Egypt, Brazil, India, Indonesia, Mexico, Nigeria, the Russian Federation, South Africa, and other great developing nations. There has even been a strong commitment to spatially balanced development in the economic history of many developed countries.*

This strong perspective against "balanced growth" in the conventional sense is important in light of the report's own assessment of evolving economic forces, in particular global integration in the era of globalization:

> *Although the basic forces shaping the internal economic geography of developing countries are the same as those that earlier shaped the economic landscapes of today's developed countries,*

> *the magnitudes have changed. Larger international markets, better transportation, and improved communication technologies mean that leading areas in open developing countries have greater market potential than industrial countries did in their early development. So the forces for spatial divergence between leading and lagging areas are now stronger.*
>
> **World Bank (2008, p. 74)**

The above perspective on openness and economic spatial disparity owes much to the burgeoning "new economic geography" literature that brings increasing returns to scale and agglomeration economies center stage in characterizing the development of an economy. In the context of a closed economy with two sectors, one in which ("agriculture") has conventional diminishing returns while the other ("manufacturing") displays firm level costs that fall as the sector as a whole grows, equilibrium can have spatial concentration of economic activity even when there is no "natural" geographic differentiation between the regions.[20] There is thus a distinction between spatial divergence caused by "first-nature geography," natural variations in environmental endowment, and "second-nature geography" that arises out of the self-enforcing feedback loops of agglomeration economies.[21]

What precisely is the impact of greater openness on spatial disparity when played through the forces of agglomeration economies? The World Bank (2008) quote above seems to suggest that spatial disparities will increase. However, the specific theory does not produce quite such a clear-cut answer. Different specifications, modeling different contexts, produce different answers.[22] For example, it matters whether different regions have equal access to the international market. It also matters whether the opening up is only for trade or also for capital mobility. The theoretical ambiguity is emphasized in recent papers by Rodríguez-Pose (2010) and Ottaviano (2009). Ottaviano (2009) summarizes the theoretical conclusions in a series of propositions as follows:

> *when regions have the same access to foreign markets, international trade liberalization fosters regional disparities and this effect is stronger the more important the foreign market and the more integrated the national market. (p. 7)...if the smaller region is a gate or a hub, international trade liberalization may reduce regional disparities. (p. 8) International capital mobility amplifies the positive effect of trade liberalization on regional disparities in the smaller country as well as in the larger one. (p. 8).*

Given these theoretical ambiguities then, what is the evidence on openness and spatial inequality? Kanbur and Venables (2007) summarize the results of a major project collating country case studies on the evolution of spatial inequality in the last quarter century. For 26 developing and transitioning countries, spatial inequality measures are available for at

[20] There is of course by now a huge literature on this. Standard references include Krugman (1991), Fujita et al. (2001), and Ottaviano and Thisse (2004).

[21] Kanbur and Venables (2007).

[22] Compare, for example, Krugman and Livas Elizondo (1996) and Paluzie (2001).

two or more points in time, so that we can get a sense of the time trends. The first and major empirical finding is that spatial inequalities have been rising in the last two to three decades.[23]

The last three decades have also been the period of globalization. Is there then a link between openness on rising spatial inequality? The case studies reported in Kanbur and Venables (2007) seem to support the hypothesis that openness is associated with greater spatial inequality. Thus, Kanbur and Zhang (2005) establish dramatic increases in spatial inequality in China since the start of the reforms in 1978. Their econometric analysis attributes at least part of this increase to the measure of openness (the other factors that are statistically significant include the degree of decentralization). Rodríguez-Pose and Sánchez-Reaza (2005) find greater regional polarization in Mexico comparing the periods before and after the North American Free Trade Agreement (NAFTA). Friedman (2005) identifies an indirect channel for Indonesia, in that openness leads to growth, but more remote areas benefit less from growth in terms of poverty reduction impact. Outside of the country studies reviewed in Kanbur and Venables (2007), Daumal (2008) finds that while for India openness contributes to greater inequality between Indian states, for Brazil the opposite is true. Thus, country context matters.

A number of cross-country regression studies have also focused on the issue of the link between openness and spatial inequality. Barrios and Strobl (2009) regress within country regional inequality against trade openness, with other controls, for 15 European Union countries. They find a positive association between regional inequality and the trade to GDP ratio for a country. Milanovic (2005) considers the evolution of regional inequality over time in China, India, the United States, Indonesia, and Brazil over 1980–2000. He finds a significant causal relationship between measures of openness and measures of regional inequality. Rodríguez-Pose and Gill (2006) analyze regional inequality similarly across country panels for the period 1970–2000. They find that it is the particular interaction of openness with the composition of trade that results in regional inequality impacts.

Perhaps the most recent and comprehensive cross-country study of regional inequality and openness is by Rodríguez-Pose (2010). It uses unbalanced panel data for 28 countries over 1975–2005. Half of these countries are developed countries and the other half are developing or transition economies. The measure of regional inequality used is the Gini coefficient of regional GDP per capita. There is no simple association between openness and regional inequality in these data. However, this is before various controls are introduced, and the panel structure of the data is exploited with appropriate techniques. On the conditioning variables, use is made of the theory referred to earlier, so that "greater trade openness will have a more polarizing effect in countries characterized

[23] Examples include Sahn and Stifel (2003) for a range of African countries, García-Verdú (2005) for Mexico, Forster et al. (2005) for Eastern Europe, Friedman (2005) for Indonesia, and Kanbur and Zhang (2005) for China.

by (a) higher differences in foreign market accessibility among its regions and (b) where there is also a high degree of coincidence between the regional income distribution and accessibility to foreign markets" (Rodríguez-Pose, 2010, p. 13). Further, like Kanbur and Zhang's (2005) work on China, it is hypothesized that the degree of decentralization will also matter for regional inequality. A number of other controls are also used, including institutional quality variables.

The overall conclusion of the comprehensive and rigorous analysis by Rodríguez-Pose (2010) is striking:

> *By and large, countries in the developing world are characterized by a series of features that are likely to potentiate the spatially polarizing effects of greater openness to trade. Their higher existing levels of regional inequality, their greater degree of sector polarization, the fact that their wealthier regions often coincide with the key entry points to trade, and their weaker state all contribute to exacerbate regional disparities as trade with the external world increases.*
>
> ***Rodríguez-Pose (2010, p. 26)***

Thus, structural differences in the country at the time of opening up tend to interact with the forces of openness, and in the recent experience at least, this has led to openness contributing to greater regional inequality. Of course, this leaves open the issue of whether this is not just the first-round effects of trade opening and whether it could it be weakened or offset by further geographical adjustments, namely, domestic migration of workers or capital, at a later stage. However, the inequality consequences in the short run will need to be addressed, and the policy implications of these findings will be discussed in a subsequent section.

20.7. INTERNATIONAL MIGRATION, REMITTANCES, AND INEQUALITY

Globalization in its most general terms is the greater integration of global economic activity. This is manifested in larger trade and in freer movement of factors of production. The vastly increased mobility of capital is often commented upon in the discourse. However, larger cross-border movement of population, from low income to high income countries, is also the subject of commentary in the popular discourse. Analytical literature has been developed to assess this phenomenon and to explore its causes and consequences. This section will provide an overview of this literature, focusing in particular on migration from developing to developed countries and on the impact of this migration on inequality in developing countries.

In 2010, the total stock of international migrants in the world (developed and developing countries) was 214 million people, up from 191 million in 2005.[24] This compares to an estimated 749 million for internal migrants. International migration is a significant

[24] International Organization for Migration (2011), p. 49.

and growing phenomenon. This is especially true of migration from developing to developed countries. The stock of immigrants in high-income countries increased at about 3% per year from 1980 to 2000. As a share of high-income country population, migrants increased from around 4% to above 8% over this 20-year period. [25]

How might the much higher rate of international migration affect the distribution of income in developing countries in theory? The answer depends on who migrates and what they do with their income after they migrate in terms of remittances to their family. If migration and remittance was representative of the domestic income distribution, then the distribution would not be affected, except for a translation to the right as remittances flowed back. Thus, poverty would decline as a result of international migration.

What if migration was not representative but selective on individual characteristics? Would the poverty results still hold? The impact effect of migration as the result of better income earning opportunities must surely be to reduce poverty at the origin. However, in the next round there is the possibility of externalities kicking in if the migrants are the most highly skilled, with knock on effects on the rest of the economy. This is the famous "brain drain" hypothesis that was popular in the 1970s and 1980s.[26] In recent years, this has been countered by the "brain gain" hypothesis, which is based on the simple idea that the probability of having access to international migration depends on the education level of the prospective migrant. In order to improve this probability, prospective migrants invest in education. Only some of these will be selected for migration, but those left behind will serve to increase the stock of human capital compared to what it would have been without the prospect of migration.[27]

There is some empirical support for the brain gain hypothesis, although others argue that its magnitude is greatly exaggerated.[28] Furthermore, there is considerable evidence for the proposition that international migration reduces poverty in the origin country. In perhaps the most comprehensive such exercise, Adams and Page (2005) asked the question on the impact of international migration on poverty using data from 71 developing countries:

> *The results show that both international migration and remittances significantly reduce the level, depth, and severity of poverty in the developing world. After instrumenting for the possible endogeneity of international migration, and controlling for various factors, results suggest that, on average, a 10% increase in the share of international migrants in a country's population will lead to a 2.1% decline in the share of people living on less than \$1.00 per person per day. After instrumenting*

[25] World Bank (2006), p. 27.
[26] See for example Bhagwati and Hamada (1974).
[27] Mountford (1997) and Stark et al. (1997).
[28] See Beine et al. (2008) for support; however, see Schiff (2005) for a skeptical perspective.

> *for the possible endogeneity of international remittances, a similar 10% increase in per capita official international remittances will lead to a 3.5% decline in the share of people living in poverty.*[29]
>
> ***Adams and Page (2005, p. 1645)***

These results are confirmed by a range of country specific studies on international migration, remittances, and poverty—examples include Acosta et al. (2006) for Latin America, Lokshin et al. (2007) for Nepal, and Adams(2006) for Ghana.

So much for poverty, where theory and evidence is relatively clear cut. What about inequality? It should be clear that selectivity of migration and remittances makes this an intricate question theoretically and empirically. And the question of identifying such selectivity is an important one in the international migration literature. In particular, there is some debate about whether migrants are selected according to education level. Using data from Docquier and Abdeslam (2006), Hanson (2010) compares the share of emigrants with tertiary education to the share of total population with tertiary education. He finds that in the vast majority of the countries, the former exceeds the latter, indicating positive selection into migration by higher levels of education. Mexico and Puerto Rico appear to be exceptions to this almost universal phenomenon, but research on migration from those origins to the United States seems to have had significant weight in the discourse. Hanson (2010) argues that a larger literature now seems to support selection on education.

What about migration selection based on unobserved variables? McKenzie et al. (2006) conduct an ingenious exercise using the results of a lottery for emigration from Tonga to New Zealand. They compare losers in the lottery with nonapplicants, both groups of course still being in Tonga. They find that the applicants have higher earnings after controlling for observables; and they conclude therefore, that those desiring to migrate are selected in terms of higher income earning potential.

If international migrants are selected from households that already have high earnings, and their migration raises income earning and, through remittances, adds to the income of the household in the origin area, it should be clear that such migration would tend to increase inequality in the sending country. However, to the extent that the selection goes the other way, inequality in the sending country will be mitigated by international migration. There is now a considerable literature on assessing directly the impact of international migration on inequality, and we now turn to an overview of those studies.

The empirical results on international migration and inequality are inconclusive as a whole. Barham and Boucher (1998) compare the actual distribution post-migration

[29] It should be noted that there is an issue in interpreting these results, which is similar to the problem of the counterfactual when microsimulating the effects of remittances. In the regression: Poverty = *f*(GDP per capita, remittances per capita) the net effect of the latter variable should be the estimated coefficient minus the (negative) change in mean income or GDP per capita due to migration times the coefficient of the mean income variable. But for this, we need an estimate of the impact of migration on the home country mean income. If it is assumed to be zero, then it is in effect assumed that migrants' labor supply is fully compensated by people remaining behind them.

including remittances for Nicaragua, with a counterfactual of what the distribution would have been if the migrants had not left and earned their original income. They found that the Gini coefficient is higher by 12%. Adams (2006) finds a much smaller increase in the Gini coefficient for Ghana—of 3%. The difference made by the counterfactual approach is illustrated by comparing the findings of De and Ratha (2005) and Karunaratne (2008) for Sri Lanka. Using the 2003–2004 Socioecononmic Survey for Sri Lanka, Karunaratne (2008) shows that "income receivers belonging to lowest 10 percent receive 1.3 percent of their income as remittances with the top 10 percent of the income receivers getting 4.6 percent of their income from remittances" (p. 58). He uses this to argue that remittances increase inequality. However, De and Ratha (2005) conduct counterfactual analysis and show that remittance income exceeds the counterfactual loss in income from migrating in the bottom two deciles, while the opposite is true for the top two deciles. Thus, they argue, remittances are equalizing.

A major issue in the empirical literature is the difference between short-term and long-term effects of international migration on inequality. In other words, the issue has to do with comparing changes in inequality in the origin location in the early stages when migration starts with when it has been going on for some time. An early study by Stark et al. (1986) found a positive relationship between remittances and inequality in the short term, but the opposite result in the long run for Mexico.[30] McKenzie and Rapoport (2007) argue that while in the short term, migration selectivity favors the better off because of the costs of migration, in the longer term these costs fall as migration networks form in the destination country. Using again the case of migration from Mexico to the United States, they argue that migration reduces inequality in communities that have experienced high levels of migration in the past. There may thus be an inverse-U relationship between international migration and inequality—first increasing and then decreasing.

Overall, then, the final effect of globalization on inequality in developing countries through the channel of international migration is ambiguous in theory, and this is reflected in the conflicting empirical findings. Of course, the migrants are themselves better off—it is the consequences for those they leave behind that are uncertain. These results pick up on a theme of this chapter as a whole, namely that the consequences for distribution depend on the context and, in particular, on preexisting structural inequalities. When these inequalities are high and interact with the opportunities presented by globalization in such a way as to benefit those who already well off, inequality will increase. The next section turns to the policy implications of these findings.[31]

[30] The analysis is based on simulating the effects of increased remittances on inequality for two villages—one of which has longstanding migration patterns to the United States and the other of which does not.

[31] There is much literature on the impact of immigration on inequalities in the host developed countries that is not covered in this chapter. For example, Borjas (2003) is the leading analyst arguing that immigration worsens inequality by lowering the relative wages of domestic low-skilled workers in the United States, while Card (2009) argues that the impact of immigration on relative wages is small, accounting for as little as 5% of the increase in U.S. age inequality between 1980 and 2000.

20.8. NATIONAL AND GLOBAL POLICY RESPONSES

Globalization brings enormous benefits, but in its wake it also brings significant risks. The risk of rising inequality has been ever present in the recent globalization discourse, where the concern has been that far from delivering "growth with equity," as it seems to have done for East Asia in the 1970s and 1980s, the more recent push to global integration has been accompanied by rising inequality. Indeed, those parts of the world that have avoided rising inequality, such as Latin America, seem to have done so through purposive policy intervention. What, then, are the policy implications of the association between globalization and rising inequality? To answer this question, bear in mind that as discussed in previous sections, our understanding of the effects of globalization on inequality, let alone of the quantitative magnitudes, is limited. This hampers policy formulation.

It helps to begin by accepting that inequality is indeed a legitimate concern for policy makers. Although not universal, there appears to be a broad consensus that rising inequality lowers social welfare directly because societies are inequality averse, and it lowers social welfare indirectly because higher inequality can impede investment and growth through a number of channels.[32] This is true of standard interpersonal inequality, as well as inequality between broadly defined groups such as gender, regions, or ethnicities. Policy makers appear to be well aware of and concerned about inequality. For example, in a survey of more than 500 Asian policy makers 44% rated concern in their country about inequality as being "high" or "very high," while 36% rated the concern as being "medium." On the question of whether higher income inequality is acceptable so long as poverty is declining, 52% disagreed or strongly disagreed. Finally, when asked how important it is to have policies in place to prevent rises in inequality in order to maintain stability and sustain growth, 95% said it was "important" or "very important."[33]

The next step in the argument is to understand that the inequality of market outcomes depends on structural inequality and on how these inequalities interact with market processes to exacerbate or mitigate these inequalities. Thus, policy can affect the inequality of final outcomes in three ways—by addressing structural inequality premarket, by addressing the operation of market processes, and by redistributing income generated by structure and market. Viewed in this way, the component parts of globalization—opening up of trade, capital, and labor flows—can be seen as dimensions of market processes. Reversing these processes in order to manage inequality is neither desirable, because it also blocks off a major route to economic growth and efficiency, nor feasible given the instruments that policy makers actually have. Of course, to the extent that the market processes are themselves distorted, for example, preferential access to foreign markets for

[32] This chapter is not the place for a review of the vast literature on this topic. A recent representative contribution is by Berg and Ostry (2011). Evidence for the detrimental effect of gender inequality and growth is presented in World Bank (2011). The effects of inequality on growth are covered in Chapter 14 of this volume.

[33] Kanbur and Zhuang (2012), p. 44.

monopolies or for politically favored groups, then addressing these can improve efficiency and equity.[34] However, policy could fruitfully focus on addressing structural inequalities and redistributing market income more equitably. Sometimes these can be combined, and redistribution of market income can be done in such a way as to mitigate structural inequalities as well.

A good entry point into policy is provided by the contrasting experiences of Asia and Latin America in the last 20 years, when both regions have faced the same global economy and increases in global integration. During the 1990s and 2000s, Asia saw sharp increases in inequality. During this period, 83% of developing Asia's population lived in countries with rising inequality, and if the high growth that occurred had taken place without rising inequality, nearly one-quarter of a billion more people would have been lifted out of poverty, according to one estimate.[35] On the other hand, Latin America, which has long been a byword for high inequality, managed to have a remarkable period of declining inequality from the late 1990s onward. This is true of all the major Latin American economies. For example, in Brazil between 1998 and 2009, without the fall in inequality, the same level of poverty reduction would have required a growth rate higher by 4 percentage points[36] Of course, the levels of inequality in Latin America were and still are much higher than those in Asia. However, the difference in trends is remarkable.

Sections 20.3 and 20.4 of this chapter discussed the skill bias that characterizes technical progress today. Demand for skilled labor is rising globally, and openness in trade and investment is transmitting this global demand to the country level. In the absence of policy intervention, these market processes will lead to rising inequality within countries. As discussed earlier, closing off economies in order to block this channel of inequality increase is neither feasible nor desirable. However, Asian economies have tended not to counteract these pressures, either by addressing structural inequalities in skill levels, or by redistributing market income sufficiently to mitigate inequality. However, Latin American economies have purposively redistributed income through cash transfers and have done it in such a way as to help the buildup of human capital through conditioning these transfers on keeping children in school. This is not the place for a full-blown assessment of conditional cash transfers (CCTs), but it does seem as though Latin American countries have found an appropriate intervention to address rising inequality in general but also for the current conjuncture of globalization-led pressures in rising inequality through a rising demand for skilled labor.[37]

[34] Inequality of assets can be compounded by inequality in market access creating inequality in the rate of return to assets for certain groups in society. This links to the discourse on inequality of opportunity, which is covered in World Bank (2005) and Chapter 4.

[35] Kanbur and Zhuang (2012), p.41.

[36] Lustig et al. (2011).

[37] For an overview, see Fiszbein and Schady (2009). It is of course clear that CCTs by themselves are not responsible for the trend of inequality in Latin America.

The additional expenditure on conditional cash transfers requires revenues, and the progressivity of the tax system is another major determinant of how globalization related increases in inequality can be mitigated. Progressivity is also important in addressing the rise in very high incomes the world over, especially in Asia. Asian tax systems do not generally score highly on progressivity. In fact, it is argued that raising progressivity of taxation would have a greater impact on inequality in Asia than elsewhere in the world.[38]

The policy discussion above is pertinent to rises in inequality associated with globalization, and it is also valid for increases in inequality from any source. What globalization brings, however, is the easier movement of capital and labor across borders, and this may well constrain government's abilities to raise revenues to address structural inequalities and to redistribute market incomes. There is now vast literature on tax competition and the globalization's role in intensifying the "race to the bottom." Kanbur and Keen (1993) show that tax rates are (i) suboptimal with lack of tax coordination when the tax base is mobile across borders and (ii) the suboptimality increases with the ease of movement of the tax base. With such revenue effects, questions are naturally raised about the sustainability of redistributive expenditure like CCTs in a globalized world. As the title of one paper asks, "Will social welfare expenditures survive tax competition?" (Hines, 2006).[39]

The basic intuitions of the analysis can be applied to progressive income taxation as well as in the context of international migration. The discussion in Section 20.7 of this chapter showed that international migration was unequivocally good for poverty reduction in developing countries, and while there were possible short-term effects raising inequality, these were turned around in the medium term. This would argue for greater freedom of international migration of labor to match the greater ease of movement of goods and capital. However, there is a catch. The possibility of international migration, especially of skilled high-income labor, could constrain the government's abilities to redistribute income within the country through progressive taxation.

The early work of Mirrlees (1982) concluded that "it may well be desirable to institute substantial income taxes on foreign earnings."[40] While this was the solution for a single country's tax design problem when faced with cross-border migration, it also contains within it the seeds of a solution to the coordination problem, whereby countries follow one another down the path of reduced progressivity, exacerbating the inequality impact of greater openness. A similar logic applies to a race to the bottom on labor standards, where countries lower standards or enforcement to gain competitive advantage (Chau and Kanbur, 2003, 2006). The issue has already been alluded to in the context of gender

[38] Asian Development Bank (2012), p. 76.

[39] See also Hines and Summers (2009).

[40] There is now vast literature on migration and optimal income taxation. A recent example that illustrates many of the intricacies is given by Hamilton and Pestieau (2005).

inequality in industries that employ mainly women. Coordination on labor standards is typically conducted through the International Labor Organization, and this mechanism can be strengthened further to address the inequality increasing forces that globalization can bring (Chau and Kanbur, 2001).

Indeed, Basu (2006) goes so far as to propose an international agency to address this issue:

> *That there may be coordination problems in trade is well recognized and we have the WTO to help mitigate such problems. That labor market policies need coordination is known and we have the ILO to address this. For environmental problems we have the UNEP or the GEF. But there is nothing comparable to these for anti-poverty and anti-inequality policies. Yet. . .this is an area where the coordination problem may be no less acute. Hence, there is clearly a perceived need for a coordinating agency. (p. 1371)*

Leaving to one side the political feasibility or operational practicality of such an agency, the fact that it is being contemplated highlights like nothing else the challenges that globalization poses to policy makers concerned with its effects on inequality.

20.9. CONCLUSION

The effects of globalization on inequality have animated much theoretical, empirical, and policy literature since World War II, but particularly so in the past 30 years when, contrary to some received wisdom, greater global integration was associated with increasing inequality in developed and especially in developing countries. In the wake of the new facts, theory has responded, particularly with a class of models that emphasizes selection mechanisms into production and trade, thereby allowing inequality to increase everywhere with openness. These new models will need to be developed, fleshed out, and applied in different contexts of trade, investment, and outsourcing. Empirical work will depend on the availability of high quality, firm-level, data, and there will need to be considerable investment in the generation of such data, particularly for low-income countries. Further, the empirical work will also need to link the firm data to household data in order to follow through on the implications for the personal distribution. More generally, there is a need to tie together the analysis of factor incomes with the implications for the personal income distribution.

Inequality is not just interpersonal inequality but also involves inequality across broadly defined groups—gender and regional groups being prime examples as well as ethno-linguistic groupings (not covered in this chapter)—adding another dimension of key policy concern. Further empirical work will need to document the impact of different aspects of globalization on these dimensions of inequality, and theorizing will need to extend and modify the standard H–O model, or indeed the more recent selection-based models, to incorporate structural divides along salient socioeconomic groupings.

At the level of national policy, addressing the inequality consequences of globalization is in principle no different than addressing the inequality consequences of other forces, such as technical progress (although global integration tightens the transmission mechanism from technical change in one part of the world to another). However, greater mobility of goods, capital, and labor constrains the freedom of governments to mitigate inequality through redistributive instruments. More research is needed to delineate, in theoretical and empirical terms, the nature of these constraints and the gains of global coordination on tax and expenditure policy and on labor and capital regulation. In the realm of practical policy, there is also a fairly full agenda, ranging from the implementation of redistributive schemes like Conditional Cash Transfers (CCTs) at the national level, and the use of existing global institutions such as the ILO and the WTO to put a floor on a race to the bottom in taxation and redistribution at the international level.

Having animated the economic analysis and policy discourse for the past half century, the globalization-inequality nexus seems set to continue doing so in the coming decades.

REFERENCES

Acosta, P., Fajnzylber, P., Lopez, H., 2006. Remittances and development in latin America. World Econ. 29 (7), 957–987.

Adams Jr., R.H., 2006. Remittances and Poverty in Ghana, World Bank Policy Research Working Paper 3838.

Adams Jr., R.H., Page, J., 2005. Do international migration and remittances reduce poverty in developing countries? World Dev. 33 (10), 1645–1669.

Adelman, I. (nd). Social development in Korea, 1953–1993. http://are.berkeley.edu/~irmaadelman/KOREA.html (accessed 19 March 2013).

Asian Development Bank, 2012. Asian Development Outlook 2012. Asian Development Bank, Manila.

Alderman, H., Chiappori, P.-A., Haddad, L., Hoddinott, J., Kanbur, R., 1995. Unitary versus collective models of the household: is it time to shift the burden of proof? World Bank Res. Obs. 10 (1), 1–19.

Atkinson, A.B., 2008. The Changing Distribution of Earnings in OECD Countries. Oxford University Press, Oxford.

Atkinson, A.B., Morelli, S., 2011. Economic Crises and Inequality, UNDP Human Development Research Paper, 2011/06.

Autor, D., 2010. The Polarization of Job Opportunities in the U.S. Labor Market: Implications for Employment and Earnings, The Center for American Progress and the Hamilton Project Working Paper (Washington, DC).

Baldacci, E., de Mello, L., Inchauste, G., 2002. Financial Crises, Poverty and Income Distribution, IMF Working Paper 02/04.

Barham, B., Boucher, S., 1998. Migration, remittances and inequality: estimating the net effects of migration on income distribution. J. Dev. Econ. 55 (3), 307–331.

Barrios, S., Strobl, E., 2009. The dynamics of regional inequalities. Reg. Sci. Urban Econ. 39 (5), 575–591.

Basu, K., 2006. Globalization, poverty and inequality: what is the relationship: what can be done? World Dev. 34 (8), 1361–1373.

Basu, P., Guariglia, A., 2007. Foreign direct investment, inequality, and growth. J. Macroecon. 29 (4), 824–839.

Beine, M., Docquier, F., Rapoport, H., 2008. Brain drain and human capital formation in developing countries: winners and losers. Econ. J. 118 (528), 631–652.
Bekaert, G., Harvey, C.R., Lundblad, C., 2006. Growth volatility and financial liberalization. J. Int. Money Financ. 25, 370–403.
Berg, A., Ostry, J.D., 2011. Equality and efficiency: is there a tradeoff between the two or do the two go hand in hand? Financ. Dev. 48 (3), 12–15.
Bhagwati, J., Desai, P., 1970. Planning for Industrialization - A Study of Indian Industrialization and Trade Policies (with J. Bhagwati). Oxford University Press, Oxford.
Bhagwati, J., Hamada, K., 1974. The brain drain, international integration of markets for professionals and unemployment: a theoretical analysis. J. Dev. Econ. 1 (1), 19–42.
Borjas, G.J., 2003. The labor demand curve is downward sloping: re-examining the impact of immigration on the labor market. Q. J. Econ. 118 (4), 1335–1374.
Bourguignon, F., Bussolo, M., 2012. Income distribution and computable general equilibrium models. In: Dixon, P.B., Jorgenson, D. (Eds.), Handbook of Computable General Equilibrium Models. Elsevier, Amsterdam. ISBN: 9780444536341.
Bourguignon, F., Silva, L.A., Bussolo, M. (Eds.), 2008. The Impact of Macroeconomic Policies on Poverty and Income Distribution: Macro–Micro Evaluation Tools. World Bank Publications, Washington, DC. ISBN: 0821357786.
Braunstein, E., Heintz, J., 2008. Gender bias and central bank policy: employment and Inflation reduction. Int. Rev. Appl. Econ. 22 (2), 173–186.
Bussolo, M., de Hoyos, R. (Eds.), 2009. Gender Aspects of the Trade and Poverty Nexus: A Micro–macro Approach. Palgrave Macmillan, London.
Chau, N., Kanbur, R., 2001. The Adoption of International Labor Standards: Who, When and Why. Brookings Trade Forum, Brookings, Washington, DC, pp. 113–156 (with N. Chau).
Chau, N., Kanbur, R., 2003. On footloose industries, asymmetric information, and wage bargaining. http://www.arts.cornell.edu/poverty/kanbur/ChauKanWageBarg.pdf.
Chau, N., Kanbur, R., 2006. The race to the bottom, from the bottom. Economica 73 (290), 193–228.
Card, D., 2009. Immigration and Inequality. NBER Working Papers, No. 14683.
Costinot, A., 2009. An elementary theory of comparative advantage. Econometrica 77, 1165–1192.
Daumal, M., 2008. Impact of trade openness on regional inequality and political unity: the cases of India and Brazil. http://www.etsg.org/ETSG2008/Papers/Daumal.pdf.
Davis, D.R., 1996. Trade Liberalization and Income Distribution. National Bureau of Economic Research, Cambridge, MA, NBER Working Paper No. 5693.
De, P., Ratha, D., 2005. Remittance Income and Household Welfare: Evidence from Sri Lanka Integrated Household Survey. Development Research Group, World Bank, Washington, DC, Unpublished paper.
Di Giovanni, J., Levchenko, A., 2008. Trade Openness and Volatility. http://www.imf.org/external/pubs/ft/wp/2008/wp08146.pdf, accessed 11 August 2012. IMF Working Paper, WP/08/146.
Docquier, F., Abdeslam, M., 2006. International migration by educational attainment, 1990–2000. In: Ozden, C., Schiff, M. (Eds.), International Migration, Remittances, and the Brain Drain. World Bank, Washington, DC, pp. 151–200, Palgrave Macmillan, New York.
Easterly, W., Islam, R., Stiglitz, J.E., 2001. Shaken and stirred: explaining growth volatility. In: Pleskovic, B., Stern, N. (Eds.), Annual World Bank Conference on Development Economics.
Feenstra, R., Hanson, G., 1996. Foreign investment, outsourcing and relative wages. In: Feenstra, R.C., et al., (Eds.), Political Economy of Trade Policy: Essays in Honor of Jagdish Bhagwati. MIT Press, Cambridge, MA, pp. 89–127.
Feenstra, R., Hanson, G., 1997. Foreign direct investment and relative wages: evidence from Mexico's maquiladoras. J. Int. Econ. 42 (3–4), 371–393.
Figini, P., Gorg, H., 1999. Multinational companies and wage inequality in the host country: the case of Ireland. Rev. World Econ. 135 (4), 594–612.
Figini, P., Gorg, H., 2011. Does foreign direct investment affect wage inequality? An empirical investigation. World Econ. 34 (9), 1455–1475.
Fiszbein, A., Schady, N., 2009. Conditional cash transfers: reducing present and future poverty: World Bank Policy Research Report. World Bank, Washington, DC.

Forster, M., Jesuit, D., Smeeding, T., 2005. Regional poverty and income inequality in central and eastern Europe: evidence form the Luxembourg income study. In: Kanbur, R., Venables, A.J. (Eds.), Spatial Inequality and Development. Oxford University Press, Oxford.
Friedman, J., 2005. How responsive is poverty to growth? A regional analysis of poverty, inequality and growth in Indonesia, 1984–99. In: Kanbur, R., Venables, A.J. (Eds.), Spatial Inequality and Development. Oxford University Press, Oxford.
Fujita, M., Krugman, P., Venables, A.J., 2001. The Spatial Economy: Cities, Regions and International Trade. MIT Press, Cambridge, MA.
García-Verdú, R., 2005. Income, mortality, and literacy distribution dynamics across states in Mexico: 1940-2000. Cuad. Econ. 42 (125), 165–192.
Goldberg, P., Pavcnik, N., 2007. Distributional Effects of Globalization in developing Countries. NBER Working Papers, No. 12885.
Grossman, G., 2013. Heterogeneous Workers and International Trade, NBER Working Paper No. 18788.
Habib, B., Narayan, A., Olivieri, S., Sanchez-Paramo, C., 2010. Assessing Ex Ante The Poverty and Distributional Impact of Global Crisis in a Developing Country, World Bank Policy Research Working Paper No. 5238.
Haddad, L., Kanbur, R., 1990. How serious is the neglect of intra-household inequality? Econ. J. 100, 866–881.
Haddad, L., Kanbur, R., 1994. Are better off households more unequal or less unequal? Oxford Econ. Pap. 46 (3), 445–458 (with L. Haddad).
Hagen, K., 2007. The changing faces of inequality in South Korea in the age of globalization. Korean Stud. 31, 1–18.
Hamilton, J., Pestieau, P., 2005. Optimal taxation and the ability distribution: implications for migration equilibria. Int. Tax Public Financ. 12, 29–45.
Hanson, G., 2010. International migration and development. In: Kanbur, R., Spence, M. (Eds.), Equity in a Globalizing World. World Bank for the Commission on Growth and Development, Washington, DC.
Helpman, E., Itskhoki, O., Redding, S., 2010. Inequality and unemployment in a global economy. Econometrica 78 (4), 1239–1283.
Hines, J.R., 2006. Will social welfare expenditures survive tax competition? Oxford Rev. Econ. Policy 22 (3), 330–348.
Hines, J., Summers, L., 2009. How Globalization Affects Tax Design. NBER Working Paper No. 14664.
Hnatkovska, V., Loayza, N., 2013. Volatility and Growth. http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2004/02/03/000012009_20040203142259/additional/105505322_20041117173010.pdf.
Hossain, N., 2011. Exports, Equity and Empowerment: The Effects of Readymade Garments Manufacturing Employment on Gender Equality in Bangladesh, Background paper for the WDR 2012.
International Organization for Migration, 2011. World Migration Report, 2011.
Irwin, D., 2008. Trade and Wages, Reconsidered: Comments and Discussion. pp. 138–143, Brookings Papers on Economic Activity, Spring.
Jomo, K., 2006. Growth With Equity in East Asia. DESA Working Paper No. 33, http://www.un.org/esa/desa/papers/2006/wp33_2006.pdf.
Juhn, C., Ujhelyi, G., Villegas-Sanchez, C., 2013. Trade liberalization and gender inequality. Am. Econ. Rev. 103 (3), 269–273.
Kabeer, N., 1997. Women, wages and intrahousehold power relations in urban Bangladesh. Dev. Change 28 (2), 261–302.
Kabeer, N., 2000. The Power to Choose: Bangladeshi Women and Labour Market Decisions in London and Dhaka. Verso, London.
Kanbur, R., 2006. The policy significance of inequality decompositions. J. Econ. Inequal. 4 (3), 367–374.
Kanbur, R., Keen, M., 1993. Tax competition and tax coordination when countries differ in size. Am. Econ. Rev. 83 (4), 877–892 (with M. Keen).
Kanbur, R., Venables, A.J., 2005. Spatial Inequality and development. In: Kanbur, R., Venables, A.J. (Eds.), Spatial Inequality and Development. Oxford University Press, Oxford, pp. 3–11 (with A.J. Venables).

Kanbur, R., Venables, A.J., 2007. Spatial disparities and economic development. In: Held, D., Kaya, A. (Eds.), Global Inequality. Polity Press, Cambridge, pp. 204–215.
Kanbur, R., Zhang, X., 2005. Fifty years of regional inequality in China: a journey through central planning, reform and openness. Rev. Dev. Econ. 9 (1), 87–106.
Kanbur, R., Zhuang, J., 2012. Confronting rising inequality in Asia. In: Asian Development Outlook 2012. Asian Development Bank, Manila.
Karunaratne, H.D., 2008. International labour migration, remittances and income inequality in a developing country: the case of Sri Lanka. Hosei Econ. Rev 75 (3), 21–65. http://repo.lib.hosei.ac.jp/bitstream/10114/1629/1/75-4hettige.pdf.
Katz, L., 2008. Trade and wages, reconsidered: comments and discussion. Brookings Pap. Econ. Act. 2008 (1), 143–149.
Kose, A., Prasad, E., Terrones, M., 2006. How do trade and financial integration affect the relationship between growth and volatility? J. Int. Econ. 69, 176–202.
Krugman, P., 1991. Increasing returns and economic geography. J. Polit. Econ. 99, 483–499.
Krugman, P., 2008. Trade and wages, reconsidered. Brookings Pap. Econ. Act. 39 (1), 103–137.
Krugman, P., Livas Elizondo, R., 1996. Trade policy and the third world metropolis. J. Dev. Econ. 49 (1), 137–150.
Lewis, W.A., 1954. Economic development with unlimited supplies of labor. Manchester Sch. Econ. Soc. Stud. 22, 139–191.
Lewis, W.A., 1976. Development and distribution. In: Cairncross, A., Puri, M. (Eds.), Employment, Income Distribution and Development Strategy: Problems of the Developing Countries (Essays in honour of H.W. Singer). Holmes & Meier Publisheres, Inc., New York, NY, pp. 26–42.
Lindert, P.H., Willamson, J.G., 2001. Does Globalization Make the World More Unequal? NBER Working Paper 8228.
Lokshin, M., Bontch-Osmolovski, M., Glinskaya, E., 2007. Work-Related Migration and Poverty Reduction in Nepal, World Bank Policy Research Working Paper.
Lustig, N., Lopez-Calva, L.F., Ortiz-Juarez, E., 2011. The Decline in Inequality in Latin America: How Much, Since When and Why. http://econ.tulane.edu/RePEc/pdf/tul1118.pdf, accessed April 20, 2013. Tulane Economics Working Paper Series, No. 1118.
McKenzie, D., Rapoport, H., 2007. Network effects and the dynamics of migration and inequality: theory and evidence from Mexico. J. Dev. Econ. 84 (1), 1–24.
McKenzie, D., Gibson, J., Stillman, S., 2006. How Important Is Selection? Experimental Versus Non-Experimental Measures of the Income Gains from Migration, World Bank Policy Research Paper No. 3906.
Melitz, M., 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. Econometrica 71 (5), 1695–1725.
Milanovic, B., 2005. Worlds Apart: Measuring International and Global Inequality. Princeton University Press, Princeton, NJ.
Mirrlees, J.A., 1982. Migration and optimal income taxes, J. Publ. Econ. 18, 319–341.
Mountford, A., 1997. Can a brain drain be good for growth in the source economy? J. Dev. Econ. 53, 287–303.
O'Donoghue, C., Loughrey, J., Morrissey, K., 2013. Using the EU-SILC to Model the Impact of Economic Crisis on Inequality, IZA Discussion paper no. 7242.
Ottaviano, G., 2009. Trade Liberalization, Economic Integration and Regional Disparities, CAF Working Papers, No. 2009/05.
Ottaviano, G., Thisse, J.-F., 2004. Agglomeration and economic geography. In: Henderson, V., Thisse, J.-F. (Eds.), In: Handbook of Regional and Urban Economics, vol. 4. Elsevier, Amsterdam.
Paluzie, E., 2001. Trade policies and regional inequalities. Pap. Reg. Sci. 80 (1), 67–85.
Papageorgiou, D., Choksi, A., Michaley, M., 1990. Liberalizing Foreign Trade in Developing Countries: The Lessons of Experience. The World Bank, Washington, DC.
Pavcnik, N., 2011. Globalization and within country inequality. In: Bacchetta, M., Jansen, M. (Eds.), Making Globalization Socially Sustainable. WTO/ILO, Geneva, pp. 233–259.
Rajan, R., 2011. Fault Lines: How Hidden Fractures Still Threaten the World Economy. Princeton University Press, Princeton.

Ravallion, M., Chen, S., 2009. The Impact of the global financial crisis on the world's poorest. http://www.voxeu.org/article/impact-global-financial-crisis-world-s-poorest.
Robilliard, A.-S., Bourguignon, F., Robinson, S., 2008. Examining the impact of the Indonesian financial crisis using a macro-micro model. In: Bourguignon, F., Silva, L.A., Bussolo, M. (Eds.), The Impact of Macroeconomic Policies on Poverty and Income Distribution: Macro-Micro Evaluation Tools. World Bank Publications, Washington, DC. ISBN: 0821357786.
Rodríguez-Pose, A., 2010. Trade and Regional Inequality, World Bank Policy Research Working Paper No. 5347.
Rodríguez-Pose, A., Gill, N., 2006. How does trade affect regional disparities? World Dev. 34, 1201–1222.
Rodríguez-Pose, A., Sánchez-Reaza, J., 2005. Economic polarization through trade: trade liberalization and regional inequality in Mexico. In: Kanbur, R., Venables, A.J. (Eds.), Spatial Inequality and Development. Oxford University Press, Oxford.
Rodrik, D., 1997. Has Globalization Gone Too Far? Institute for International Economics, Washington, DC 1997.
Sahn, D., Stifel, D., 2003. Urban-rural inequality in living standards in Africa. J. Afr. Econ. 12 (1), 564–597.
Schiff, M., 2005. Brain Gain: Claims About Its Size and Impact on Welfare and Growth are Greatly Exaggerated, IZA Discussion Paper No. 1599.
Seguino, S., 2007. Is more mobility good? Firm mobility and the low-wage low productivity trap. Struct. Change Econ. Dyn. 18 (1), 27–51.
Seguino, S., 2013. From micro-level gender relations to the macro economy and back again: theory and policy. In: Figart, D., Warnecke, T. (Eds.), Handbook of Research on Gender and Economic Life. Edward Elgar, Northampton, MA.
Stark, O., Edward Taylor, J., Yitzhaki, S., 1986. Remittances and inequality. Econ. J. 96 (383), 722–740.
Stark, O., Helmenstein, C., Prskawetz, A., 1997. Abrain gain with a brain drain. Econ. Lett. 55, 227–234.
Takhtamanova, Y., Sierminska, E., 2009. Gender, monetary policy, and employment: the case of nine OECD countries. Feminist Econ. 15 (3), 323–353.
Taylor, K., Driffield, N., 2005. Wage inequality and the role of multinationals: evidence from UK panel data. Labour Econ. 12 (2), 223–249.
Tejani, S., Milberg, W., 2010. Global Defeminization? Industrial Upgrading, Occupational Segregation, and Manufacturing, SCEPA Working Paper 2010-1.
Tsai, P.L., 1995. Foreign direct investment and income inequality: further evidence. World Dev. 23 (3), 469–483.
Van Staveren, I., Elson, D., Grown, C., Cagatay, N. (Eds.), 2007. The Feminist Economics of Trade. Routledge, London.
Verhoogen, E., 2008. Trade, quality upgrading and wage inequality in the Mexican manufacturing sector. Q. J. Econ. 123 (2), 489–530.
Wood, A., 1994. North-South Trade, Employment and Inequality: Changing Fortunes in a Skill-Driven World. Clarendon Press, Oxford.
Wood, A., 1999. Openness and wage inequality in developing countries: the Latin American challenge to East Asian conventional wisdom. In: Baldwin, R., et al. (Eds.), Market Integration, Regionalism and the Global Economy. Cambridge University Press, Cambridge.
World Bank, 1993. The East Asian Miracle: Economic Growth and Public Policy. World Bank, Washington, DC.
World Bank, 2005. World Development Report 2006: Equity and Development. World Bank, Washington, DC.
World Bank, 2006. Global Economic Prospects 2006. World Bank, Washington, DC.
World Bank, 2008. World Development Report 2009: Reshaping Economic Geography. World Bank, Washington, DC.
World Bank, 2011. World Development Report 2012: Gender and Development. World Bank, Washington, DC.
Zhang, X., Kanbur, R., 2001. What difference do polarisation measures make? An application to China. J. Dev. Stud. 37 (2001), 85–98.

CHAPTER 21

Democracy, Redistribution, and Inequality

Daron Acemoglu*, **Suresh Naidu**[†], **Pascual Restrepo***, **James A. Robinson**[‡]
*Massachusetts Institute of Technology, Cambridge, MA, USA
[†]Columbia University, NY, USA
[‡]Harvard University, Cambridge, MA, USA

Contents

Handbook of Income Distribution, Volume 2B
ISSN 1574-0056, http://dx.doi.org/10.1016/B978-0-444-59429-7.00022-4

Abstract

In this paper we revisit the relationship between democracy, redistribution, and inequality. We first explain the theoretical reasons why democracy is expected to increase redistribution and reduce inequality, and why this expectation may fail to be realized when democracy is captured by the richer segments of the population; when it caters to the preferences of the middle class; or when it opens up disequalizing opportunities to segments of the population previously excluded from such activities, thus exacerbating inequality among a large part of the population. We then survey the existing empirical literature, which is both voluminous and full of contradictory results. We provide new and systematic reduced-form evidence on the dynamic impact of democracy on various outcomes. Our findings indicate that there is a significant and robust effect of democracy on tax revenues as a fraction of GDP, but no robust impact on inequality. We also find that democracy is associated with an increase in secondary schooling and a more rapid structural transformation. Finally, we provide some evidence suggesting that inequality tends to increase after democratization when the economy has already undergone significant structural transformation, when land inequality is high, and when the gap between the middle class and the poor is small. All of these are broadly consistent with a view that is different from the traditional median voter model of democratic redistribution: democracy does not lead to a uniform decline in post-tax inequality, but can result in changes in fiscal redistribution and economic structure that have ambiguous effects on inequality.

Keywords

Democracy, Education, Inequality, Political development, Redistribution, Structural transformation

JEL Classification Codes

P16, O10

21.1. INTRODUCTION

Many factors influence the distribution of assets and income that a market economy generates. These include the distribution of innate abilities and property rights, the nature of technology, and the market structures that determine investment opportunities and the distribution of human and physical capital.

But any market system is embedded in a larger political system. The impact of the political system on distribution depends on the laws, institutions, and policies enacted by that system. What institutions or policies a political system generates depends on the distribution of power in society and how political institutions and mobilized interests aggregate preferences. For example, we expect institutions that concentrate political power within a narrow segment of the population—typical of nondemocratic regimes—to generate greater inequality.[1]

[1] Nondemocracies tend to be dominated by the rich either because the rich wield sufficient power to create such a regime or because those who can wield power for other reasons subsequently use this power to become rich.

As the literature has shown, there are several theoretical mechanisms through which such an impact might operate. One would be the enactment of policies benefiting the politically powerful at the expense of the rest of society, including policies pushing down wages by repression and other means. In Apartheid South Africa prior to 1994, for example, the political system dominated by the minority white population introduced government regulations on the occupation and residential choices of black Africans in order to reduce their wages (e.g., by reducing competition for white labor and by forcing blacks into unskilled occupations, see Lundahl, 1982; Wilse-Samson, 2013). Another mechanism is the one highlighted by Meltzer and Richard's (1981) seminal paper. Building on earlier research by Romer (1975) and Roberts (1977), they developed a model where extensions of the voting franchise, by shifting the median voter toward poorer segments of society, increase redistribution, and reduce inequality.[2]

Despite these strong priors, the empirical literature is very far from a consensus on the relationship between democracy, redistribution, and inequality. Several works have reported a negative relationship between democracy and inequality using specific historical episodes or cross-national studies. Acemoglu and Robinson (2000) argued this was the case based on the economic history of nineteenth-century Europe and some twentieth-century Latin American examples. An important study by Rodrik (1999) presented evidence from a panel of countries that democracy is associated with higher real wages and higher labor share in national income. Lindert (1994, 2004) provided evidence from OECD countries indicating a linkage between democratization and public spending, particularly on education; Persson and Tabellini (2003) presented similar cross-national evidence; and Lapp (2004) pointed to a statistical association between democratization and land reform in Latin America. Other papers point in the opposite direction, however. Sirowy and Inkeles (1990) and Gradstein and Milanovic (2004) have argued that the cross-national empirical evidence on democracy and inequality is ambiguous and not robust. Scheve and Stasavage (2009, 2010, 2012) have claimed that there is little impact of democracy on inequality and policy among OECD countries, and Gil et al. (2004) have forcefully argued that there is no relationship between democracy and any policy outcome in a cross section of countries (Perotti, 1996, was an earlier important paper with similar negative findings).

In this chapter we revisit these issues in a unified theoretical and empirical framework. Theoretically, we review the standard Meltzer-Richard model and point out why the relationship between democracy, redistribution, and inequality may be more complex than the standard model might suggest. First, democracy may be "captured" or "constrained." In particular, even though democracy clearly changes the distribution of de jure power in society (e.g., Acemoglu and Robinson, 2006), policy outcomes

[2] Historically, the fear of expected redistribution has been one of the factors motivating the opposition to democracy (see Guttsman, 1967).

and inequality depend not just on the de jure but also the de facto distribution of power. For example, Acemoglu and Robinson (2008) argue that, under certain circumstances, those who see their de jure power eroded by democratization may sufficiently increase their investments in de facto power (e.g., via control of local law enforcement, mobilization of nonstate armed actors, lobbying, and other means of capturing the party system) in order to continue to control the political process. If so, we would not see an impact of democratization on redistribution and inequality.[3] Similarly, democracy may be constrained by either other de jure institutions such as constitutions, conservative political parties, and judiciaries, or by de facto threats of coups, capital flight, or widespread tax evasion by the elite.

Second, we suggest that democratization can result in "inequality-increasing market opportunities." Nondemocracy may exclude a large fraction of the population from productive occupations (e.g., skilled occupations) and entrepreneurship (including lucrative contracts) as in apartheid South Africa or the former Soviet bloc countries. To the extent that there is significant heterogeneity within this population, the freedom to take part in economic activities on a more level playing field with the previous elite may actually increase inequality within the excluded or repressed group and consequently the entire society.[4]

Finally, consistent with Stigler's (1970) "Director's law", democracy may transfer political power to the middle class rather than to the poor. If so, redistribution may increase and inequality may be curtailed only when the middle class is in favor of such redistribution.

After reviewing the fairly large and heterogeneous prior literature on this topic, the rest of this chapter examines the empirical impact of democracy on tax revenues as a percentage of GDP (as an imperfect measure of redistribution) and on inequality as well as a number of additional macro variables. We evaluate previous empirical claims about the effect of democracy in a consistent empirical framework that controls for a number of confounding variables. Our objective is not to estimate some structural parameters or the "causal" effect of democracy on redistribution, but to uncover whether there is a

[3] Relatedly, there could be reasons for dictators to redistribute and reduce inequality to increase the stability of that regime (e.g., Acemoglu and Robinson, 2001; Albertus and Menaldo, 2012, more generally). Plausible cases of this would be the land reform implemented by the Shah of Iran during his White Revolution of 1963 to help him become more autonomous from elites (McDaniel, 1991), the agrarian reforms made by the Peruvian military regime in the early 1970s (chapter 2 of Seligmann, 1995), or the educational reforms in 19th-century oligarchic Argentina (Elis, 2011).

[4] Our data show that inequality has in fact increased in South Africa between 1990 and 2000 (or 2005) and in ex-Soviet countries between 1989 and 1995 (or 2000), periods that bracket their democratic transitions in 1994 and 1989 respectively. This is probably, at least in part, driven by the increase in inequality among previously disenfranchised blacks and repressed citizens (for details on the post-democracy distributions of income see Whiteford and Van Seventer, 2000, for South Africa and Milanovic, 1998, for ex-Soviet countries).

robust correlation between democracy and redistribution or inequality, and to undertake a preliminary investigation of how this empirical relationship changes depending on the stage of development and various other factors potentially influencing how democracy operates.

The previous literature has used several different approaches (e.g., cross-sectional regressions, time-series and panel data investigations) and several different measures of democracy. We believe that cross-sectional (cross-national) regressions and regressions that do not control for country fixed effects will be heavily confounded with other factors likely to be simultaneously correlated with democracy and inequality. We therefore focus on a consistent panel of countries, and investigate whether countries that become democratic redistributed more and reduced inequality relative to others. We also focus on a consistent definition of democracy based on Freedom House and Polity indices, building on the work by Papaioannou and Siourounis (2008). One of the problems of these indices is the significant measurement error, which creates spurious movements in democracy. To minimize the influence of such measurement error, we create a dichotomous measure of democracy using information from both the Freedom House and Polity datasets as well as other codings of democracy to resolve ambiguous cases. This leads to a measure of democracy covering 184 countries annually from 1960 (or post-1960 year of independence) to 2010. We also pay special attention to modeling the dynamics of our outcomes of interest, taxes as a percentage of GDP, and various measures of structural change and inequality.

Our empirical investigation uncovers a number of interesting patterns (why many of these results differ from some of the existing papers in the literature is discussed after they are presented). First, we find a robust and quantitatively large positive effect of democracy on tax revenue as a percentage of GDP (and also on total government revenues as a percentage of GDP). The long-run effect of democracy in our preferred specification is about a 16% increase in tax revenues as a fraction of GDP. This pattern is robust to various different econometric techniques and to the inclusion of other potential determinants of taxes, such as unrest, war, and education.

Second, we find a positive effect of democracy on secondary school enrollment and the extent of structural transformation (e.g., an impact on the nonagricultural share of employment and the nonagricultural share of output).

Third, however, we find a much more limited effect of democracy on inequality. In particular, even though some measures and some specifications indicate that inequality declines after democratization, there is no robust pattern in the data (certainly nothing comparable to the results on taxes and government revenue). This may reflect the poorer quality of inequality data. But we also suspect it may be related to the more complex, nuanced theoretical relationships between democracy and inequality pointed out above.

Fourth, we investigate whether there are heterogeneous effects of democracy on taxes and inequality consistent with these more nuanced theoretical relationships. The evidence here points to an inequality-increasing impact of democracy in societies with a high degree of land inequality, which we interpret as evidence of (partial) capture of democratic decision making by landed elites. We also find that inequality increases following a democratization in relatively nonagricultural societies, and also when the extent of disequalizing economic activities is greater in the global economy as measured by U.S. top income shares (though this effect is less robust). These correlations are consistent with the inequality-inducing effects of access to market opportunities created by democracy. We further find that democracy tends to increase inequality and taxation when the middle class is less prosperous relative to the poor. These correlations are consistent with Director's law, which suggests that democracy often empowers the middle class to redistribute from the rest of society to itself. Our results suggest the need for a more systematic investigation of the conditions under which democracy does indeed reduce inequality and increase redistribution.

The chapter proceeds as follows. In the next section we discuss the theoretical connections between democracy, redistribution, and inequality. In Section 21.3 we provide a survey of the existing empirical literature on the impact of democracy on taxes, redistribution, inequality, and some other reduced-form dependent variables potentially associated with inequality (e.g., average calories per person, life expectancy, and infant mortality). Section 21.4 then describes our econometric methodology and data. Section 21.5 presents our new findings, and Section 21.6 concludes.

21.2. THEORETICAL CONSIDERATIONS

In this section, we illustrate some of the linkages between democracy and inequality that have been proposed in the literature. We begin with the seminal Meltzer and Richard (1981) model, but then alter the set of instruments available to the government to show how the logic of the standard model can be altered and even reversed. We will discuss the impact of democracy, modeled as a broader franchise, relative to a nondemocratic regime modeled as a narrower franchise or controlled by a small group. This broadening of access to political power is what our primary cross-country empirical measures of democracy attempt to capture, and is arguably the most important feature of a democratic regime.

21.2.1 The Redistributive and Equalizing Effects of Democracy

We start with the standard "equalizing effect" of democracy, first emphasized formally in Meltzer and Richard's (1981) seminal study (see also Acemoglu and Robinson, 2006). Democratization, by extending political power to poorer segments of society, will increase the tendency for pro-poor policy naturally associated with redistribution, and thus reduce inequality.

Suppose that society consists of agents distinguished only with respect to their endowment of income, denoted by y_i for agent i, with the distribution of income in the society denoted by the function $F(y)$ and its mean by $\overline{y}$. The only policy instrument is a linear tax τ imposed on all agents, with the proceeds distributed lump-sum again to all agents. We normalize total population to 1 without loss of any generality.

The government budget constraint, which determines this lump-sum transfer T, takes the form

$$T \leq \tau\overline{y} - C(\tau)\overline{y}, \tag{21.1}$$

where the second term captures the distortionary costs of taxation. $C(\tau)$ is assumed to be differentiable, convex and nondecreasing, with $C'(0) = 0$.

Each agent's post-tax income and utility is given by

$$\hat{y}_i = (1-\tau)y_i + \tau\overline{y} - C(\tau)\overline{y}. \tag{21.2}$$

This expression immediately makes it clear that preferences over policy—represented by the linear tax rate τ—satisfy both single crossing and single-peakedness (e.g., Austen-Smith and Banks, 1999). Hence the median voter theorem, and its variants for more limited franchises (see e.g., Acemoglu et al., 2012) hold.[5]

Suppose, to start with, that there is a limited franchise such that all agents with income above y_q, the q^{th} percentile of the income distribution, are enfranchised and the rest are disenfranchised. Consider a "democratization," which takes the form of y_q decreasing, say to some $y_{q'} < y_q$, so that more people are allowed to vote. Let the equilibrium tax rate under these two different political institutions be denoted by τ_q and $\tau_{q'}$, and the resulting post-tax income distribution by F_q and $F_{q'}$. Then from the observation that the median of the distribution truncated at $y_{q'}$ is always less than the median for the one truncated above $y_q > y_{q'}$, the following result is immediate:

Proposition 1
Redistributive Effects of Democracy
Suppose that starting from only those above y_q being enfranchised, there is a further democratization so that now those above $y_{q'} < y_q$ are enfranchised. This democratization leads to higher taxes ($\tau_{q'} \geq \tau_q$), higher redistribution, and a more equal distribution of post-tax income in the sense that $F_{q'}$ is more concentrated around its mean than F_q.

A few comments about this proposition are useful. First, this result is just a restatement of Meltzer and Richard's (1981) main result. Second, the first part of the conclusion is stated as $\tau_{q'} \geq \tau_q$, since if both y_q and $y_{q'}$ are above the mean, with standard arguments, $\tau_{q'} = \tau_q = 0$. Third, the second part of the conclusion does *not* state that F_q is a

[5] Namely, if we assume that policy choices are made by either a direct democracy procedure choosing the Condorcet winner (if one exists) or as a result of competition between two parties choosing (and committing to) their platforms, the equilibrium will coincide with the political bliss point of the median-ranked voter. As Austen-Smith and Banks (1999) discuss in detail, these types of results, though powerful, are rather special and rely, among other things, on the assumption that the policy space is unidimensional.

mean-preserving spread of, or is second-order stochastically dominated by $F_{q'}$, because higher taxes may reduce mean post-tax income due to their distortionary costs of taxation. Instead, the statement is that $F_{q'}$ is more concentrated around its mean than F_q, which implies the following: if we shift $F_{q'}$ so that it has the same mean as F_q, then it second-order stochastically dominates F_q (and thus automatically implies that standard deviation and other measures of inequality are lower under $F_{q'}$ than under F_q).

Finally, the result in the proposition should be carefully distinguished from another often-stated (but not unambiguous) result, which concerns the impact of inequality on redistribution. Persson and Tabellini (1994) and Alesina and Rodrik (1994), among others, show that, under some additional assumptions, greater inequality leads to more redistribution in the median voter setup (which in these papers is also embedded in a growth model). This result, however, is generally not true.[6] It applies under additional assumptions on the distribution of income, such as a log normal distribution, or when the gap between mean and median is used as a measure of inequality (which is rather nonstandard). In contrast, the result emphasized here is unambiguously true.

This result of Meltzer and Richard (1981) is the basis for the hypothesis that democracy should increase taxation and income redistribution and reduce inequality. In the model, the only way that redistribution can take place is via a lump-sum transfer. This is obviously restrictive. For example, it could be that individuals prefer the state to provide public goods (Lizzeri and Persico, 2004) or public education. Nevertheless, the result generalizes, under suitable assumptions, to the cases in which the redistribution takes place through public goods or education.

We next discuss another possible impact of democracy and why its influence on redistribution and inequality may be more complex than this result may suggest.

21.2.2 Democracy and the Structural Transformation

The logic of Proposition 1 applies when the main political conflict involves the tax rate but not other policy instruments. One of the most important alternatives, emphasized by Moore (1966) and by Acemoglu and Robinson (2006) in the economics literature, is the combination of policies used to create abundant (and cheap) labor for the rural sector (see also Llavador and Oxoby, 2005). Many nondemocratic agrarian societies use explicit and implicit limits on migration out of the rural sector, together with labor repression, to keep wages low and redistribute income from the population to the politically powerful landed elites. Even industrial sectors in nineteenth century England used the Master and Servant

[6] Consider the following counterexample. In society A, 1/3 of the population has income 2, 1/3 has income 3 and the remaining 1/3 has income 7. If everyone is enfranchised, the Condorcet winner is a tax rate $\tau^A > 0$ with $C'(\tau^A) = 1/4$. In society B, 1/3 of the population has income 0, 1/3 has income 4 and the remaining 1/3 has income 8. If everyone is enfranchised, the Condorcet winner is a tax rate $\tau^B = 0$. Society B has a lower tax rate, and hence less redistribution despite being more unequal (the distribution of income in society A second-order stochastically dominates the distribution of society B).

law to prosecute workers and repress trade unions, and it was only repealed following an expansion of the franchise to workers and decriminalization of workers' organizations (Naidu and Yuchtman, 2013). For example, in rural Africa, land is often controlled by traditional rulers and chiefs and not held as private property. People moving away from particular chieftaincies lose rights over land, which inhibits migration. In Sierra Leone, forced labor controlled by chiefs was common in rural areas prior to the civil war in 1991 (e.g., Acemoglu et al., 2014). We may expect that these policies will be relaxed or lifted when political power shifts either to industrialists, who would benefit from migration out of the rural sector into the industrial one, or to poorer segments of society who are bearing the brunt of lower wages (see Acemoglu, 2006, for a political economy analysis of wage repression and the impact of democracy on it).

To model these issues in the simplest possible way, suppose that there is a single policy instrument denoted by $\eta \in \mathbb{R}_+$ capturing the extent of barriers against mobility out of the rural sector. Suppose now that y_i denotes the land endowment of agent i, so that post-policy income (and utility) of an agent is given by

$$\hat{y}_i = \omega(\eta) + \upsilon(\eta) y_i, \tag{21.3}$$

where $\omega(\eta)$ can be interpreted as the impact of this policy on wage income (thus it applies agents with no land endowment) and naturally we assume that $\omega(\eta)$ is decreasing. On the other hand, $\upsilon(\eta)$ is the impact of its policy on land rents, and is thus increasing. This formulation can also be easily extended to include industrialists who may also be opposed to high values of η, which would reduce the supply of labor to their sector.

Inspection of Equation (21.3) immediately reveals that preferences over η satisfy single crossing, and thus the median voter theorem again applies. This leads to the following result:

Proposition 2
Democracy and Structural Transformation
Consider the model outlined in this subsection. Suppose that starting from only those above y_q being enfranchised, there is a further democratization such that now those above $y_{q'} < y_q$ are enfranchised. This democratization leads to lower mobility barriers out of the rural sector ($\eta_{q'} \le \eta_q$) and a more equal distribution of income (in the sense that $F_{q'}$ is more concentrated around its means than F_q).

This proposition highlights that the same reasoning that leads to the redistributive and equalizing effects of democracy also weighs in favor of lifting barriers that are against the interest of the middle class and the poor. An important implication of this might be a push toward the structural transformation out of agriculture and into industry and cities that might have been partly arrested artificially by the political process before democratization. An illustrative example of this is the impact of the 1832 Reform Act in Britain, which enfranchised urban manufacturing elites in the newly industrializing cities such as Birmingham and Manchester. This led directly to the abolition of the Corn Laws in 1846 which was a huge distortionary subsidy to landowners (Schonhardt-Bailey, 2006).

It is also straightforward to apply this reasoning to other policies related to redistribution and structural transformation, such as investment in mass schooling, which we may also expect to be boosted by democratization.

21.2.3 Other Considerations

Obviously, the simple model presented in the previous two subsections leaves out many mechanisms which might influence the extent of redistribution in a democracy and other forces that can shape the political equilibrium (Putterman, 1996, provides an overview of many ideas).[7]

Several papers have investigated how social mobility influences the demand for redistribution even in a democracy (Alesina and La Ferrara, 2005; Bénabou and Ok, 2001; Carter and Morrow, 2012; Wright, 1996). When rates of social mobility are high and tax policy is sticky, people who are poor today may not support high rates of taxation and redistribution because they worry that it will negatively impact them should they become rich in the future. Relatedly, Piketty (1995) suggests that different beliefs about distortionary taxation can be self-fulfilling and lead to multiple equilibria, some with low inequality and a lot of redistribution, and others with high inequality and little redistribution (see also Alesina and Angeletos, 2005; Bénabou, 2001, 2008; Bénabou and Tirole, 2006). Thus, a democratic society could result in an equilibrium with little redistribution.

Alternatively, it could be that social cleavages or identities may be such as to reduce the likelihood that a coalition favoring redistribution would form (De la O and Rodden, 2008; Frank, 2005; Lee, 2003; Roemer, 1998; Roemer et al., 2007; Shayo, 2009). For example, in Roemer's model there is a right-wing political party that does not like taxation and redistribution and a left-wing political party that does. People are ideologically predisposed toward one of the parties, but they also care about religion, as do the parties. If the right-wing party is Catholic, a poor Catholic may vote for it even if it does not offer the tax policy that the voter wishes. Another reason that the above model may fail to characterize the political equilibrium accurately is because ethnic heterogeneity limits the demand for redistribution (Alesina and Glaeser, 2004; Alesina et al., 1999). Daalgard et al. (2005) argue that institutions, particularly ones that influence the

[7] We have also left out a discussion of several other important issues that have been raised in theoretical analysis of redistribution in democracy. In particular, there is a growing and vibrant literature on redistribution in a dynamic context, including Krusell et al. (1997), Krusell and Ríos-Rull (1999), Hassler et al. (2003), Battaglini and Coate (2008), and Acemoglu et al. (2012). Overviews of other aspects of democratic policymaking are provided in Drazen (2000), Persson and Tabellini (2000), Acemoglu and Robinson (2006), and Besley (2007). The political economy literature on the emergence of democracy is also beyond the scope of our chapter, and we refer the reader to the extensive discussions in Acemoglu and Robinson (2006).

efficiency of the state, will influence the demand for redistribution. Finally, recent work has tied the amount of social capital to the extent of redistribution such as in Scandinavia (Algan et al., 2013).

Another idea, due to Moene and Wallerstein (2001), is that most redistribution under democracy does not take the form of transfers from rich to poor but of social insurance. Moene and Wallerstein develop a model to show that the comparative statics of this with respect to inequality may be very different from the Meltzer-Richard model.

In the rest of this section, we will instead focus on what we view as the first-order mechanisms via which democracy may fail to increase redistribution or reduce inequality.

21.2.4 Why Inequality May Not Decline: Captured Democracy and Constraints on Redistribution

In contrast to Propositions 1 and 2, greater democratization may not always reduce inequality. In this and the next two subsections, we discuss several mechanisms for this.

The first possible reason is that even though democracy reallocates de jure power to poorer agents, richer segments of society can take other actions to offset this by increasing their de facto power. This possibility, first raised in Acemoglu and Robinson (2008), can be captured in the following simple way here. Suppose that the distribution of income has mass at two points, the rich elite, who are initially enfranchised, and the rest of the citizens, who make up the majority of the population and are initially disenfranchised. Suppose, in addition, that the rich elite can undertake costly investments to increase their de facto power (meaning the power they control outside those that are strictly institutionally sanctioned, such as their influence on parties' platforms via lobbying or repression through control of local law enforcement or nonstate armed actors; see Acemoglu and Robinson, 2006, 2008; Acemoglu et al., 2013b,c). If they do so, they will "capture the political system," for example, control the political agenda of all parties or change political ideology via the media. Suppose also that this type of capture is costly, with cost denoted by $\Gamma > 0$. Then clearly, when there is a limited franchise, the elite will not need to incur the cost for doing so. Once there is enfranchisement, if this cost is not too large, they will find it beneficial to incur this cost, and may then succeed in setting the tax rate at their bliss point, rather than putting up with the higher redistribution that the majority of citizens would impose.

This reasoning immediately implies the following result:

Proposition 3
Captured Democracy
Suppose that the elite can control the political system after democratization at cost $\Gamma > 0$. Then if Γ is less than some $\overline{\Gamma}$, they will prefer to do so, and democratization will lead to no change in taxes and the distribution of income.

This proposition, in a simple way, captures the main idea of Acemoglu and Robinson (2008), even though the specific mechanism for capture is somewhat different. In Acemoglu and Robinson, each elite agent individually contributes to their collective de facto power, which needs to be greater in democracy to exceed the increased de jure power of poor citizens. Under some conditions, the main result of Acemoglu and Robinson (2008) is that the probability of the elite controlling political power is invariant to democratization—or more generally may not increase as much as it may have been expected to do owing to the direct effect of the change in de jure power.

A related channel to Proposition 3 is that democracy may be highly dysfunctional, or effectively captured, because its institutional architecture is often chosen by previous restricted franchises or dictatorships. Acemoglu et al. (2011) develop a model where the elite can take control of democracy by forming a coalition in favor of the continuation of patronage, keeping the state weak.

Other mechanisms include de jure constitutional provisions that restrict the scope for redistribution (e.g., a cap on τ) after democratization. For instance, Siavelis (2000) and Londregan (2000) argue that the constitution imposed by the Pinochet government in Chile prior to the transition to democracy was a way to constrain future redistribution. Another is the threat of a future coup preventing democracy from pursuing high redistribution. Ellman and Wantchekon (2000) discuss how fear of a military coup induced voters to support the right-wing ARENA party, taking redistribution off the political agenda, and also suggest that similar forces operated in electing Charles Taylor in Liberia in 1997 (see also Acemoglu and Robinson, 2001). An alternative mechanism is the threat of capital flight increasing the cost of redistribution (in the reduced-form model here, this would mean an increase in $C(\tau)$).[8] Moses (1994) argues that this was the case for Sweden in 1992, as well as Campello (2011) and Weyland (2004), among others, who suggest that capital flight restrained redistribution in new Latin American democracies (see also Acemoglu and Robinson, 2006). Mohamed and Finnoff (2003) similarly argue that capital flight constrained redistribution in post-apartheid South Africa (see also Alesina and Tabellini, 1989; Bardhan et al., 2006). All of these constraints would reduce the potential impact of democracy on inequality.

An implication of Proposition 3 and our discussion is that democracy may change neither fiscal policy nor the distribution of income. Nevertheless, it is also useful to note that a variant of this model can lead to an increase in taxes without a major impact on inequality. Suppose, for example, that the elite can use their de facto power to redirect spending toward themselves (e.g., toward some public goods that mostly benefit the elite such as investments in elite universities rather than in primary or secondary education),

[8] A related idea, proposed by Dunning (2008), is that if the main source of tax revenues is from natural resource rents, rather than personal income or wealth taxes, the elite have less incentive to oppose or capture democracy.

but have a more limited ability to control taxes. In that case, a variant of Proposition 3 would apply whereby democracy might be associated with an increase in taxation, but may not have a major impact on inequality. Moreover, in the Acemoglu et al. model mentioned above, democracy may increase taxes in order to use them as payments to state employees, but still not increase redistribution or reduce inequality.

Another variant of this result where elites can block democratization ex-ante, rather than capturing democracies ex-post, shows how selection bias can affect the correlation between democracy and the extent of redistribution observed. If elites can block democratizations that would be highly redistributive, then the only democratizations that are observed would be those that are not particularly redistributive, and we would see no correlation between democracies and increased taxation or redistribution.

A number of studies present empirical evidence consistent with these mechanisms. Larcinese (2011), for example, shows that the democratization of Italy in 1912, though it had a large positive effect on the number of people who voted, had little impact on which parties were represented in the legislature, something he interprets as consistent with the democracy being captured by old elites. Berlinski and Dewan (2011) similarly show that the British Second Reform Act of 1868, though it greatly expanded voting rights, did not have a significant immediate impact on representation.

Anderson et al. (2011) show that in Maharashtra in Western India, areas where the traditional Maratha landlords are powerful as measured by their landholdings, have democratic equilibria that are far more pro-landlord and anti-poor because the Maratha elites control voting behavior via their clientelistic ties to workers. See also Baland and Robinson (2008, 2012) on Chile; McMillan and Zoido (2004) on Peru; Pettersson-Lidbom and Tyrefors (2011) on Sweden; and Albertus and Menaldo (2014) for a cross-country empirical study of how the strength of elites at the time of democratization influences how redistributive democracy is.

There is also qualitative historical evidence on the redistributive constraints faced by democracies. Writers since James Madison have argued that the U.S. constitution is an effective bulwark against redistribution (Beard, 1913; Holton, 2008; McGuire, 2003). Others have noted that the constitution was a large obstacle to slave emancipation (Einhorn, 2006; Waldstreicher, 2009), and Dasgupta (2013) argues that the Indian constitution has been a key component in elites maintaining control of land reform projects.

21.2.5 Why Inequality May Not Decline: Inequality-Increasing Market Opportunities

Our second mechanism for an ambiguous effect of democracy on inequality is inspired by the experiences of South Africa and Eastern Europe. In South Africa, the end of apartheid in 1994 has been associated with an increase in inequality. This is partly because the black majority now takes part in economic activities from which it was previously excluded, and earnings are more dispersed in these activities than the low-skill, manual occupations to which they were previously confined. Likewise in Eastern Europe after 1989, the collapse of communism created new opportunities for people who were previously trapped in sectors of the economy where they could not use their skills and talents optimally (Atkinson and Micklewright, 1992; Flemming and Micklewright, 2000).

To incorporate this possibility, let us return to the model of structural transformation presented above. Suppose that y_i denotes the "skill" endowment of agent i, and is strictly positive for all agents. Now $\eta \in \{0,1\}$ denotes a policy instrument preventing people from moving into some potentially high-productivity activity, with $\eta = 1$ representing such prevention and $\eta = 0$ as its cessation. Post-policy income of agent i is

$$\hat{y}_i = v(\eta) y_i \mathbf{I}\left(y_i > y_q\right) + (1-\eta) y_i + w_0,$$

where $v(\eta)$ denotes the return to agents above the $q^{\text{th}} > 0.5$ percentile of the distribution (e.g., the landowners) from preventing the rest of the population's entrance into the high-productivity activities (e.g., banning black workers in South Africa from skilled occupations). The indicator function $\mathbf{I}(y_i > y_q)$ makes sure that this term only applies to agents above the q^{th} percentile. In view of this, it is natural to assume that $v(\eta = 1) > v(\eta = 0) + 1$ so that the very rich benefit from this policy. In addition, if $\eta = 1$, then the remaining workers just receive a baseline wage $w_0 > 0$. In contrast, if $\eta = 0$, they are able to take part in economic activities, and in this case, some of them, depending on their type, will be more successful than others.

The median voter theorem still applies in this formulation, and following democratization extending the franchise sufficiently, the political process will lead to a switch to $\eta = 0$. However, this formulation also makes it clear that the increased market opportunities for agents below the q^{th} percentile will create inequality among them. This effect can easily dominate the reduction in inequality resulting from the fact that the very rich no longer benefit from restricting access for the rest of the population. We summarize this result in the next proposition:

Proposition 4
Implications of Inequality-Inducing Market Opportunities
In the model described in this subsection, suppose there is an increase in democracy. If a sufficient number of voters are enfranchised, this will lead to a switch from $\eta = 1$ to $\eta = 0$, but the implications for inequality are ambiguous.

21.2.6 Why Inequality May Not Decline: The Middle Class Bias

The third possible reason for a limited impact of democracy on inequality is that, with additional tax instruments, greater democratization may empower the middle class (loosely and broadly defined), which can then use its greater power to redistribute to itself. Suppose society now consists of three groups: the rich elite with income y_r, the middle class with income $y_m < y_r$, and the poor with income $y_p < y_m$. Let the proportions of these three groups be, respectively, δ_r, δ_m, and δ_p. Consider an extension of the baseline model where there are two types of transfers: the lump-sum transfer, T, as before, and a transfer specifically benefiting the middle class, denoted by T_m. The government budget constraint is then

$$T + \delta_m T_m \leq \tau\bar{y} - C(\tau)\bar{y}. \tag{21.4}$$

Now suppose that starting with the rich elite in power there is a democratization, which makes the median voter an agent from the middle class. This will be the case if there is a limited franchise extension only to the middle class and $\delta_r < \delta_m$ (the middle classes are more populous than the rich), or there is a transition to full democracy but the middle class contains the median voter (i.e., $\delta_r + \delta_p < \delta_m$). Clearly, when only the elite are empowered there will be zero taxation (because, given the available fiscal instruments, the elite cannot redistribute to itself). With the middle class in power, there will be positive taxation and redistribution to the middle class using the instrument T_m. The resulting income distribution may be more or less equal (it will be more equal if the middle class is much poorer than the rich, **and less equal if the middle classes are much richer than the poor**).

In this case, the impact of democracy on inequality is generally ambiguous and depends on the specific measure of inequality under consideration, the cost of taxation and the pre-democracy distribution of income. It can be shown that, focusing on the Gini coefficient, when the poor are numerous and not too poor relative to the rich, that is, when

$$\frac{\delta_p}{1-\delta_p} y_p > \frac{\delta_r}{1-\delta_r} y_r, \tag{21.5}$$

inequality increases under democracy.[9] Intuitively, in this case, taxes hurt the poor who also do not benefit from the transfers. When the poor are more numerous and richer, they bear more of the burden of taxation, and this can increase inequality.

Furthermore, whether democratization increases or reduces inequality depends on the shares of income accruing to the rich and the poor before democracy. When either

[9] In particular, the Gini coefficient under autocracy is
$G^A = \delta_p - \delta_r + s_r(\delta_m + \delta_r) - s_p(\delta_p + \delta_m)$,
where the s's denote the income shares of the rich and the poor. The Gini coefficient under democracy can be computed with the same formula but using the post-tax income shares of the rich and the poor, e.g., $\hat{s}_g = s_g(1-\tau^D)/(1-C(\tau^D))$, as
$G^D = \delta_p - \delta_r + s_r \frac{1-\tau^D}{1-C(\tau^D)}(\delta_m + \delta_r) - s_p \frac{1-\tau^D}{1-C(\tau^D)}(\delta_p + \delta_m)$.
The change in the Gini due to democratization is then
$G^D - G^A = s_p\left(\frac{\tau^D - C(\tau^D)}{1-C(\tau^D)}\right)(\delta_p + \delta_m) - s_r\left(\frac{\tau^D - C(\tau^D)}{1-C(\tau^D)}\right)(\delta_m + \delta_r)$.
Noting that $\tau^D > C(\tau^D)$, the result follows.

Equation (21.5) holds or when C is sufficiently convex that the tax choice of the middle class is not very elastic, an increase in the share of income of the rich or a decrease in the share of income of the poor makes it more likely that democracy will reduce inequality.[10] These results are summarized in the next proposition.

Proposition 5
Modified Director's Law
In the model described in this subsection, suppose there is limited enfranchisement to the middle class and $\delta_r < \delta_m$, or there is a transition to full democracy and $\delta_r + \delta_p < \delta_m$. Then there will be an increase in taxes but the effect on inequality—measured by the Gini coefficient—is ambiguous. If Equation (21.5) holds, democracy increases the Gini coefficient. Moreover, if either Equation (21.5) does not hold or C is sufficiently convex, then a larger share of income of the rich (which always increases taxes) makes it more likely that inequality will decline under democracy. If either Equation (21.5) holds or C is sufficiently convex, then a larger share of income of the poor (which also always increases taxes) makes it more likely that inequality will increase under democracy.

We refer to this result as the "Modified Director's law" since it relates to an idea attributed to Aaron Director by Stigler (1970) that redistribution in democracy involves taking from the poor and the rich to the benefit of the middle class (one can derive a similar result in a model of probabilistic voting when the middle class has a larger density for the distribution of its valence term, Persson and Tabellini, 2000, section 7.4).

This result is also related to what Aidt et al. (2009) call the "retrenchment effect" of democratization. They show that local franchise expansion in nineteenth-century Britain to the middle class often reduced expenditure on public good provision since the middle class bore the brunt of property taxes which financed local public good provision. In their model, an expansion of voting rights, by reducing public good provision and taxes on the

[10] First note that higher shares of income of the rich and the poor always increase the preferred tax rate of the middle class $\frac{d\tau^D}{ds_r} > 0$ and $\frac{d\tau^D}{ds_p} > 0$. Next, following on from Footnote 9, the impact of the share of income of the rich on the change in the Gini is
$\frac{d}{ds_r}(G^D - G^A) = -H(\tau^D)(\delta_m + \delta_r) + \left[s_p\left(\delta_p + \delta_m\right) - s_r(\delta_m + \delta_r)\right]H'^{D})\frac{d\tau^D}{ds_r}$,
where $H(\tau) = (\tau - C(\tau))/(1 - C(\tau))$ is the share of revenue taken by the government in taxes, which is increasing provided that $C'(\tau), C(\tau) < 1$, and $\tau > C(\tau)$, which are automatically satisfied when τ is to the right of the peak of the Laffer curve. The first term, corresponding to the incidence of taxation on the rich, is always negative. The second term is also negative when Equation (21.5) does not hold (otherwise higher taxes, creating more resources to be transferred to the middle class, are dis-equalizing), or dominated by the first term when $\frac{d\tau^D}{ds_r} > 0$ is small, which is the case when C is sufficiently convex (so that taxes do not respond significantly to an increase in s_r).
Similarly, the impact of the share of income of the poor on the changing Gini is given by
$\frac{d}{ds_p}(G^D - G^A) = H(\tau^D)\left(\delta_p + \delta_m\right) + \left[s_p\left(\delta_p + \delta_m\right) - s_r(\delta_m + \delta_r)\right]H'^{D})\frac{d\tau^D}{ds_p}$.
The first term is now positive because inequality increases when the poor bear more of the tax burden. The second effect is also positive when Equation (21.5) holds, or dominated by the first term when C is sufficiently convex.

middle class, can thus increase inequality. Relatedly, Fernandez and Rogerson (1995) show how an equilibrium like this could arise in a political economy model of taxation and educational subsidies.

An important contrast between this result and Proposition 3 is on taxes. In Proposition 3, democracy neither increases taxes nor reduces inequality (but note the contrast with extended versions of the captured democracy mechanism). Here democracy increases taxes, but because the additional revenue is used for the middle class, it may not reduce inequality.[11]

21.2.7 Discussion and Interpretation

The theoretical ideas presented so far suggest that in the most basic framework, we expect democracy to increase redistribution and reduce inequality. We may also expect a boost to structural transformation from democratization. However, several factors militate against this tendency. The elite—the richer segments of society—who stand to lose from increased redistribution can attempt to increase their de facto power to compensate for their reduced de jure power under democracy. As we have seen, this can limit redistribution and/or the potential reduction in inequality. Alternatively, consistent with Director's law, democracy may indeed increase taxes but use the resulting revenues for redistribution to the middle class, thus not necessarily reducing inequality. Finally, democracy may also be associated with the opening up of new economic opportunities to a large segment of society, which can be an additional source of inequality.

After reviewing the existing empirical literature, we will investigate the impact of democracy on redistribution and inequality. We will, in particular, study whether the effect of democracy on redistribution and inequality is heterogeneous and whether it depends on the economic and political forces we have highlighted in this section. In line with the theoretical mechanisms here, we expect the captured democracy effect to be stronger if the elite have more to lose from democracy, for example, if they are more vested in land or other assets that will lose value when wages increase and nondemocratic policies useful for these assets are lifted. Additionally, we expect the position of the middle class in the distribution of income to shape the type and extent of redistribution observed in democracy. Finally, we also expect the inequality-inducing market opportunity effect to be stronger when frontier technologies and global economic activities are more human or physical capital-biased and when society is more urbanized and presents greater opportunities for entrepreneurship and capitalist development. These are some of the ideas we will investigate in greater detail in the empirical analysis.

[11] While we do not explore this in the chapter, this result also suggests that measures of polarization, as discussed in Chapter 5, could be an important source of heterogeneity in the relationship between democracy and redistribution, as the middle class would have more to gain from taxing both the poor and the rich.

21.3. PREVIOUS LITERATURE

In this section, we survey the literature on the effect of democracy on redistribution and inequality. Our emphasis will be on the empirical literature, though we also discuss some of the theoretical ideas that have played an important role in this literature (several theoretical contributions have already been discussed in the previous section).

21.3.1 Democracy, Taxes, and Redistribution

In the basic model of the policy effects of democracy proposed by Meltzer and Richard (1981), an expansion of democracy should lead to greater tax revenues and redistribution. We first consider the tax and spending part of this. While Gil et al. (2004) found no correlation between tax revenues and different components of government spending and democracy in a cross-sectional specification, as we discuss below, there are many studies which do find such results.

This is certainly true of the more historical studies, for example, Lindert (2004), Gradstein and Justman (1999a), and Acemoglu and Robinson (2000). Aidt et al. (2006) and Aidt and Jensen (2009b) examine the impact of democratization measured by the proportion of adults who could vote in a cross-national panel consisting of 12 Western European countries over the period 1830–1938, and in a sample of 10 Western countries over the period 1860–1938, respectively. The latter paper, for example, finds robust positive effects of suffrage on government expenditure as a percentage of GDP and also tax revenues as a percentage of GDP.

One would expect that democracy not only changes the total amount of tax revenues, but also what taxes were used for. For instance, one might expect democracies to move towards more progressive taxation. Aidt and Jensen (2009b) investigated the impact of suffrage on tax incidence. They found, somewhat paradoxically, that suffrage expansion led to lower direct taxes and higher indirect taxes. Aidt and Jensen (2009a) investigated the determinants of the introduction of an income tax. They reported a nonlinear relationship with suffrage, indicating that an expansion of the franchise starting from very restrictive levels reduces the probability that an income tax will be introduced, but also that this probability increases significantly at higher levels of the franchise.

Scheve and Stasavage (2010, 2012) also adopt a long-run approach using data from OECD countries and find no correlation between democracy and either tax progressivity or the rate of capital taxation. Instead, consistent with Tilly (1985) and Besley and Pearson (2011), they emphasize the importance of warfare, a topic to which we return later.

An important study by Lindert (1994) found an impact of democracy on various types of social spending in a panel data consisting of European and North American countries as well as Japan, Australasia, Argentina, Brazil, and Mexico and spanning the period from 1880 to 1930. In his 2004 book, Lindert summarizes his findings as: "Conclusion #1: There was so little social spending of any kind before the twentieth century mainly because political voice was so restricted" (Lindert, 2004, p. 22).

A lot of research is consistent with this. Huber and Stephens (2012) build a panel dataset for Latin America between 1970 and 2007 and measure democracy by the cumulative years a country has been democratic since 1945 and estimate pooled OLS models without fixed effects. They find the history of democracy is significantly positively correlated with education spending, health spending and Social Security, and welfare spending. In a panel data of 14 Latin American countries for 1973–1997, Kaufman and Segura-Ubiergo (2001) show that democracy, as measured by the dichotomous measure introduced by Przeworski et al. (2000), is positively correlated with government expenditure on health and education but not with other components of spending. Brown and Hunter (1999) also focus on Latin America using a panel between 1980 and 1992. They examine the impact of democracy, coded as a dichotomous measure based on Przeworski et al. (2000), on social spending per capita. They also examine various types of interactions between democracy and other variables such as GDP per capita and the growth rate in GDP per capita. Their basic findings suggest that democracies have greater social spending than autocracies.

Using a broader set of countries and a panel between 1960 and 1998, Persson and Tabellini (2003) also find some evidence that democracy, as measured by the Gastil index and the Polity score, has positive effects on government expenditure and government revenues as well as welfare and Social Security spending as percentages of GDP.

Though most studies tend to focus on a broad measure of democracy, an interesting literature has examined female enfranchisement more specifically. The main focus of this research has been on whether enfranchising women has an additional or differential impact on government taxation or spending. Lindert (1994) showed that female enfranchisement had an independent effect on social spending and this finding has held up well (see Aidt and Dallal, 2008, for similar results for a later period). Lott and Kenny (1999) studied the expansion of women's voting rights in the United States between 1870 and 1940 and found that it coincided with increases in per capita state revenues and expenditures. Miller (2008) also examined this process showing that female suffrage increased health spending and led to significant falls in infant mortality.

Of all the research on this topic, only the paper by Aidt and Jensen (2013) provides an identification strategy to tackle the fact that democracy is endogenous. Building on the theoretical ideas in Acemoglu and Robinson (2000, 2006) and their previous work (Aidt and Jensen, 2011), they argue that "revolutionary threat," measured by revolutionary events in other countries, is a viable instrument for democracy in a panel of Western European countries between 1820 and 1913. Using this source of variation, they find that democracy, as measured by the extent of suffrage (proportion of the adult population that is enfranchised), has a robust positive effect on government spending relative to GDP.

In this light, the paper by Gil et al. (2004) appears an outlier in finding no effects of democracy on tax revenues as a percentage of GDP and spending. Nevertheless, there are econometric problems with all of these papers. Specifically, there is little attention to identification problems and most studies that use panel data do not include country fixed effects, thus confounding the effect of democracy with country-specific factors potentially correlated with democracy and redistribution. Though the important study of Aidt and Jensen (2013) moves the literature a long way forward, their empirical model controls for many endogenous variables on the right side and does not deal with the possibility that revolutionary events in other countries might capture other correlated effects impacting the outcomes of interest (see the discussion of this possibility in Acemoglu et al., 2013a).

21.3.2 Democracy and Inequality

There is an even larger reduced-form empirical literature on the relationship between democracy and inequality, most of it by sociologists and political scientists rather than economists. This has typically delivered ambiguous results. Early work, which consisted mostly of simple cross-national regressions of measures of inequality (usually the income Gini coefficient) on various measures of democracy, was surveyed by Sirowy and Inkeles (1990). They concluded "the existing evidence suggests that the level of political democracy as measured at one point in time tends not to be widely associated with lower levels of income inequality" (p. 151).

Much of this literature, however, also suffers from the econometric problems of the type discussed in the last subsection. Most importantly, there is the possibility that omitted factors are affecting both inequality and democracy, and that reverse causation from inequality to democracy may be present (e.g., Muller, 1988).

Muller (1988), using a larger dataset than the previous literature, found that there was a negative correlation between the number of years a country had been democratic and inequality, which he interpreted as evidence that democracy had to be in place for long enough for inequality to fall. Yet the robustness of his results were challenged by Weede (1989) (see the response by Muller, 1989). Others, such as Simpson (1990), Burkhart (1997), and Gradstein and Justman (1999b) claimed that there was a nonlinear reduced-form relationship between democracy and inequality with inequality being low at both low and high levels of democracy and higher for intermediate levels. The plethora of results is what led Sirowy and Inkeles to be skeptical, though they do suggest that there may be some evidence in favor of the relevance of the history of democracy for inequality (Muller's original finding has been replicated in many subsequent studies, e.g., by Huber et al., 2006; Huber and Stephens, 2012, table 5.10). Nevertheless, there are good reasons for being skeptical about these findings, since the impact of the history of democracy is identified in models that do not include fixed effects, and obviously, it will capture the impact of these omitted fixed effects. More generally, this is just a special case of the difficulty of identifying duration dependence and unobserved heterogeneity—a difficulty that this literature neither tackles nor recognizes.

Three more recent studies used better data and exploited the time as well as the cross-sectional dimensions to investigate the impact of democracy on inequality. Rodrik (1999) showed that either the Freedom House of Polity III measure of democracy was positively correlated with average real wages in manufacturing and the share of wages in national income (in specifications that also control for productivity, GDP per capita and a price index). He illustrated this both in a cross section and in a panel of countries using country fixed effects. He also presented evidence that political competition and participation at large were important parts of the mechanisms via which democracy worked.[12] Scheve and Stasavage (2009) used a long-run panel from 1916 to 2000 for 13 OECD countries with country fixed effects and found that universal suffrage, measured as a dummy, had no impact on the share of national income accruing to the top 1%. Perhaps consistent with a variant of the (upper) middle class bias argument we provided above, they found that there is actually a statistically significant positive correlation between the universal suffrage dummy and what they called the "Top10-1" share, which is the share of income accruing to people between the 90th and 99th percentiles of the income distribution divided by the share accruing to the people above the 99th percentile. Finally, Li et al. (1998) used pooled OLS to show that an index of civil liberties is negatively correlated with inequality (greater civil liberties, lower inequality) though they do not investigate the relationship between inequality and more conventional measures of democracy.

Though this research has been dominated by studies that examine the average effect of democracy, Lee (2005) uses a panel data random effects model to argue that there are heterogeneous effects of democracy on inequality. The panel is unbalanced and covers 64 countries between 1970 and 1994. In particular, he argues that there is a significant interaction between the size of government as measured by tax revenues as a percentage of GDP and democracy. The paper finds that, although there is a significant positive correlation between democracy and inequality, the interaction between democracy and the size of government is significant and negative, suggesting that for large enough levels of government, democracy reduces inequality. Lee interprets this as measuring state strength (similarly to Cheibub, 1998 and Soifer, 2013).

21.3.3 Education and Democracy

The impact of democracy on education has also been examined both historically and using contemporary cross-national data and some of the results were noted in the last section. The work of Lindert (2004, chapter 5) is again central and, as with his work on social spending, Lindert presents evidence that the historical emergence of democracy is connected with educational expansion. A complementary historical study by Engerman and Sokoloff (2005, 2011) points out that within the Americas there is a close

[12] We will return to Rodrik's study below, and particularly in Appendix A, to explain the contrast between his and our results.

connection between the extent of democracy, measured by voting rights, the proportion of adults that voted and an effective secret ballot, and measures of education such as literacy rates.

A great deal of econometric work supports this research using various measures of education. Baum and Lake (2001), for example, found that secondary-school gross enrollment rates also increased with democracy across the developing world, "particularly among regimes that have experienced large changes in democracy" (p. 613) (see also Baum and Lake, 2003). Brown and Hunter (2004), focusing on 17 Latin American countries between 1980 and 1997, find that the Polity index is positively correlated with total educational expenditures per capita and also with the share of expenditures going into primary education. This finding mirrors the earlier one of Brown (1999) who finds that various dichotomous measures of democracy created from the Polity dataset and the measure of Przeworski et al. (2000) were positively correlated with primary school enrollment. Huber and Stephens (2012) also find robust evidence in Latin America for a positive correlation between the history of democracy and educational spending (see also Avelino et al., 2005).

These issues have also been intensively studied in sub-Saharan Africa. Stasavage (2005a) examined the impact of democratization in the 1990s in Africa on education, using a measure of democracy similar to Przeworski et al. (2000), and presented evidence that democracy increases total educational spending as a percentage of GDP. He also found evidence of increases in spending on primary education as a percentage of GDP, though this was not robust to the use of country fixed effects. Stasavage (2005b) provides a case study of democratization and educational expansion in Uganda. More recent research by Harding and Stasavage (2013) reconfirms the impact of democracy on primary education, this time looking at primary enrollment, and shows that the likely channel runs through a greater probability that democratic governments will abolish primary school fees.

Gallego (2010) presents one of the few attempts to develop an identification strategy to examine the impact of democracy on education. There are many reasons why this is important. Most obviously, there is the issue of whether or not there is reverse causation from education to democracy. Though the results of Acemoglu et al. (2005) reduce this concern, the above papers deal with this at best by using lagged democracy as an explanatory variable. Gallego follows Acemoglu et al. (2001, 2002) and uses their data on the historical settler mortality of Europeans and indigenous population density in 1500 as instruments for democracy and finds that democracy in 1900, measured by the Polity score, has a significant causal effect on primary school enrollment in 1900. Gallego recognizes that the exclusion restriction of his instrument may be violated but provides a very careful discussion of the potential biases that this involves and how this works against the findings he focuses on, arguing that he estimates a lower bound on the effect of democracy on education.

Using a broad sample of over 100 countries between 1960 and 2000, Ansell (2010) uses panel data regressions with and without country fixed effects to examine the impact of democracy, measured by the Polity score, on various components of educational spending. He also instruments for democracy using lagged democracy and the levels of democracy in neighboring countries. He finds that democracy has a positive and significant effect on total educational spending as a percentage of GDP, and on educational spending as a percentage of the government budget. Using cross-national regressions he also finds a negative correlation between democracy and private educational spending as a percentage of GDP and also between democracy and primary school expenditure per student by the government. He argues, contrary to Stasavage, that democracy tilts educational spending away from primary and toward secondary and tertiary education.

The likely reconciliation of all these results is that the type of education democracy produces depends on what forces democracy unleashes and who wields power in democracy. In Uganda, when President Museveni allowed democratization, he did so in a society lacking a large middle class who could dominate educational spending decisions. Hence as Stasavage showed, primary school enrollment increased. But in a large cross-national sample, the relationship may be dominated by dictatorships that spend more on primary schooling and democracies that focus on secondary schooling (see also Gradstein et al., 2004; Ansell, 2010, for relevant models).

This may also account for the results in recent work by Aghion et al. (2012), which uses a long but unbalanced panel of 137 countries between 1830 and 2001 and reports a negative correlation between the Polity score and primary school enrollment.

21.3.4 Democracy and Health Outcomes

There is also some other work on the impact of democracy on health outcomes. These are potentially related to inequality, because rapid improvements in health outcomes tend to come at the bottom of the distribution. Many studies, for example, find that democracy is positively correlated with life expectancy (see McGuire, 2010, for an overview and case study and econometric evidence). Besley and Kudamatsu (2006) show this in a panel data model for the post-war period but without using country fixed effects. Wigley and Akkoyunlu-Wigley (2011) in a complementary study have shown that life expectancy is positively correlated with the history of democracy of a country. Kudamatsu (2012) showed in the context of democratic transitions in Africa that health outcomes improved in countries that democratized compared to those that did not.

Blaydes and Kayser (2011) looked at the relationship between democracy and average calories per capita interpreted as a proxy for inequality, because calories consumed decline very quickly with income. Using a trichotomous measure of democracy based on the Polity IV dataset (where greater than 7 is a democracy, less than -7 is an autocracy, and everything in between a "hybrid regime"), they show in a panel data model with country fixed effects that democracy is positively correlated with average calorie consumption.

Gerring et al. (2012) find using panel data from 1960 to 2000 that, although the current level of democracy, as measured by the Polity score, is not robustly correlated with infant mortality, there is a strong negative correlation between the history of democracy and infant mortality—the more a country has experienced democracy in the past, the lower is infant mortality currently. Contrary to these findings, Ross (2006), using panel data from 1970 to 2000, the Polity score, the Przeworski et al. (2000) dichotomous measure of democracy, and the history of democracy as independent variables, finds no robust correlation between any of them and infant and child mortality. A possible reconciliation of these findings is that, as mentioned above, the history of democracy is nothing but a proxy for the omitted fixed effects, and Ross obtains different results from Gerring et al. because he controlled for fixed effects. Another confounding factor is that this literature in general does not control for the dynamics of democracy and GDP per capita and the endogeneity of democratization (see Acemoglu et al., 2013).

21.3.5 The Intensive Margin

All the papers discussed so far use various national-level measures of democracy, usually based on well-known databases created by political scientists. An important complementary direction is to investigate within-country variation exploiting other measures of "effective" enfranchisement.

In this context, particularly interesting is Fujiwara's (2011) study of changes in the voting technology in Brazil in the 1990s. These, by making it much simpler and easier for illiterate people to vote, massively enfranchised the poor. Fujiwara estimates the effect of this change by exploiting differences in the way the policy was rolled out. He shows that the consequence of the reform was a change in government spending in a pro-poor direction, particularly with respect to health expenditures, and that infant mortality fell as a result. Baland and Robinson (2008, 2012) examine another related reform, the introduction of an effective secret ballot in Chile in 1958. Though they do not directly study any policy outcomes, they do show that the reform led to large increases in the vote share of left-wing parties, which, they argue, is consistent with this democratizing reform moving the political equilibrium towards more pro-poor policies. They also find that land prices fall, which illustrates that the price of land capitalized the value of controlling workers' votes under the open ballot.

Martinez-Bravo et al. (2012) study the effects of elections in China on redistribution and public good provision. They use variation in the introduction of village elections in China, controlling for village and year fixed effects as well as province-level trends. They find that village chairmen experience higher turnover and become more educated and less likely to be Communist Party members following the introduction of elections. They also find that taxes and public goods increase as a result of the elections. In particular, irrigation increases more in villages with more farmland, and public education increases in villages with more children. They also find that income inequality is reduced, and less land is leased to elite-controlled enterprises.

Naidu (2011) examined the impact of the disenfranchisement of blacks in the US South via poll taxes and literacy tests in the period after the end of Reconstruction. He finds that this reversal of democracy reduced the teacher-student ratio in black schools by 10–23%, with no significant effects on white teacher–student ratios. Also, consistent with Baland and Robinson's results, disenfranchisement increased farm values.

Relatedly, using state-level data Husted and Kenny (1997) examine the impact of the abolition of literacy tests and poll taxes in the United States over the period 1950–1988 and find that this was associated with a significant increase in welfare expenditures but not other types of government expenditures. Using county-level data, Cascio and Washington (2012) find that expansion of voting rights in the South resulted in increased state transfers to previously disenfranchised counties. Besley et al. (2010), on the other hand, show that the abolition of literacy tests and poll taxes was associated with increased political competition in US states. Increased political competition between the Republicans and Democrats reduced government tax revenues relative to state income and increased infrastructure expenditure relative to other components of government expenditure.

21.4. ECONOMETRIC SPECIFICATION AND DATA

Given the conflicting results in the theoretical and empirical literature surveyed above, we now present our econometric framework for investigating the relationship between democracy, redistribution, and inequality. We attempt to evaluate the diverse results within a single empirical strategy and sample, and we provide what we view to be some basic robust facts.

In this section, we describe our econometric specifications and our main data. Our approach is to estimate a canonical panel data model with country fixed effects and time effects while also modeling the dynamics of inequality and redistribution. Both fixed effects and allowing for dynamics (e.g., mean reversion) are important. Without fixed effects, as already noted above, several confounding factors will make the association between democracy and inequality (or redistribution) difficult to interpret. Moreover, we will see that there are potentially important dynamics in the key outcome variables, and failure to control for this would lead to spurious relationships (or make it difficult to establish robust patterns even when such patterns do exist).

Some of the papers we mentioned above have adopted a set-up similar to this, for example Rodrik (1999), Ross (2006), Scheve and Stasavage (2009), Aghion et al. (2012), and Aidt and Jensen (2013), but without modeling the dynamics in inequality or redistribution. In addition, several of these papers suffer from the "bad control" problem; for example, Scheve and Stasavage (2009) control for both suffrage and education in their investigation of the determinants of the top income shares. If democracy influences inequality via its impact on education, then such an empirical model is bound to find that democracy is not correlated with inequality. Even the pioneering paper by Aidt and Jensen (2013) controls for many endogenous variables on the right side of the regression including the Polity score of the country.[13]

21.4.1 Econometric Specification

Consider the following simple econometric model:

$$z_{it} = \rho z_{it-1} + \gamma d_{it-1} + \mathbf{x}'_{it-1}\beta + \mu_t + \psi_i + u_{it}, \tag{21.6}$$

where z_{it} is the outcome of interest, which will be either (log of) tax revenue as a percentage of GDP or total revenue as a percentage of GDP as alternative measures of taxation, education, structural change, or one of several possible measures of inequality. The dependent variables with significant skewness in their cross-country distribution, in particular, tax to GDP ratio, total government revenues to GDP ratio, agricultural shares of employment, and income and secondary enrollment, will be in logs, which makes interpretation easier and allows the impact of democracy to be proportional to the baseline level. All of the results emphasized in this paper also hold in specifications using levels rather than logs, but these are not reported to conserve space. Lags in this specification will always mean 5-year lags: d_{it-1} is democracy 5 years ago. The lagged value of the dependent variable on the right-hand side is included to capture persistence (and mean reversion) in these outcome measures, which may be a determinant of democracy or correlated with other variables that predict democracy. The main right hand side variable is d_{it}, a dummy for democracy in country i in period t whose construction will be described in detail below. This variable is lagged by one period (generally a 5-year interval) because we expect its impact not to be contemporaneous. All other potential covariates, as well as interaction effects which are included later, are in the vector $\mathbf{x}_{it-1}$, which is lagged to avoid putting endogenous variables on the right-hand side of the regression. In our baseline specification, we include lagged log GDP per capita as a covariate for several reasons.[14] First, as we show in Acemoglu et al. (2013), democracy is much more likely to suffer from endogeneity concerns when the lagged effects of GDP per capita are not controlled for. Second, in Acemoglu et al. (2013), we also show that democracy has a

[13] A more desirable approach would be to develop an instrument for democracy. We believe that the only credible papers on this topic are Gallego (2010), Aidt and Jensen (2013), and our own work, Acemoglu et al. (2013). We do not pursue these directions as this would take us too far from our purpose of surveying and interpreting the literature and presenting what we believe to be the robust correlations in the data.

[14] We will always use GDP to refer to log GDP per capita.

major effect on GDP per capita and changes in GDP per capita may impact inequality independently of the influence of democracy on this variable. In all cases, we also report specifications that do not control for GDP per capita to ensure that the results we report are not driven by the presence of this endogenous control.

Finally, the ψ_i's denote a full set of country dummies and the μ_t's denote a full set of time effects that capture common shocks and trends for all countries. u_{it} is an error term, capturing all other omitted factors, with $E[u_{it}|z_{it-1}, d_{it-1}, \mathbf{x}'_{it-1}, \mu_t, \psi_i] = 0$ for all i and t. We estimate the above equation excluding the Soviet Union and its satellite countries because the dynamics of inequality and taxation following the fall of the Soviet Union are probably different from other democratizations. In some cases, for example, when using the tax to GDP ratio, this restriction is irrelevant because there is no data for these countries. When there is data, as with inequality, we also report results including these countries.

Our estimation framework controls for two key sources of potential bias. First, it controls for country fixed effects, which take into account that democracies are different from nondemocracies in many permanent characteristics that we do not observe and that may also affect inequality and taxation.[15] Second, it allows for mean-reverting dynamics and persistent effects in the dependent variable that may be endogenous to democracy.[16] This focus on *changes* in democracy ignores variation across countries that never change political institutions, for example, the United States, India, and China, but these observations help us in forming the counterfactual outcome conditional on the right-hand side covariates. Put differently, countries that never change political institutions may still be informative about how taxation and inequality change as a function of past taxation and inequality.

The simplest way of estimating Equation (21.6) is by OLS and imposing $\rho = 0$, and this is the most common regression in the prior literature which has used panel data. But, as already pointed out above, if $\rho > 0$, this specification may lead to biased estimates and will not correctly identify the long-run effect of democracy on the outcome of interest. An alternative method is to estimate this equation by OLS (which is just the standard within-group estimator removing the fixed effects by eliminating the mean of country i). This estimator is not consistent when the number of time periods is finite, because the regressor z_{it-1} is mechanically correlated with u_{is} for $s < t$, and this will induce a downward bias in the estimate of ρ (e.g., Wooldridge, 2002, chapter 11). However, the bias

[15] For instance, democracies may have more pluralistic institutions or stronger states, which may independently affect inequality and taxation.

[16] For instance, crisis, turmoil, social unrest, or increases in inequality could trigger a democratization, and also have a persistent effect on the path of our dependent variable. In this case, it becomes important to control for the dynamics of taxes or inequality by adding their lag on the right-hand side.

becomes smaller as the number of periods grows, holding ρ constant, so for large enough T or low enough ρ it becomes negligible (Nickell, 1981).

Our preferred estimation strategy is to deal with this econometric problem using a standard generalized method of moments (GMM) estimator along the lines of Holtz-Eakin et al. (1988) and Arellano and Bond (1991). This involves differencing Equation (21.6) with respect to time

$$\Delta z_{it} = \rho \Delta z_{it-1} + \gamma \Delta d_{it-1} + \Delta \mathbf{x}'_{it-1}\beta + \Delta \mu_t + \Delta u_{it}, \tag{21.7}$$

where the fixed-country effects are removed by time differencing. Although Equation (21.7) cannot be estimated consistently by OLS either, in the absence of serial correlation in the original residual, u_{it} (i.e., no second-order serial correlation in Δu_{it}), z_{it-2} and all further lags, and thus also d_{it-2} and all further lags, are uncorrelated with Δu_{it}, and can be used as instruments for Δz_{it-1}, incorporating them as moment conditions in a GMM procedure.

An alternative procedure removes country fixed effects by taking forward orthogonal differences. In particular, for variable w_{it}, this is given by

$$w_{it}^{\text{fod}} = \sqrt{\frac{T_{it}}{T_{it+1}}}\left(w_{it} - \frac{1}{T}\sum_{s>t} w_{is}\right),$$

where T_{it} is the number of times w_{is} appears in the data for $s > t$. Forward orthogonal differences also remove the fixed effects. In the absence of serial correlation in the original residual, z_{it-1}, d_{it-1}, $\mathbf{x}'_{it-1}$ and all further lags are orthogonal to the transformed error term u_{it}^{fod}, and can be used to form moment conditions in a GMM procedure. Moreover, if the original residuals were i.i.d., then the transformed error term will also be i.i.d.[17]

We will implement this using Arellano and Bond's GMM estimator with different subsets of moments, and after taking first differences or forward orthogonal differences of the data. As Newey and Windmeijer (2009) show, using the full set of moments in two-step GMM may lead to the "too many instruments" bias, since the number of potential moments one could use to estimate the dynamic panel model is quadratic in the time dimension. Thus, we experiment by restricting the number of lags used to form moments in the estimation. In addition to restricting the number of moments, we focus on

[17] Estimates of the model obtained by taking forward orthogonal differences are different from the first difference estimates only in unbalanced panels or when not all Arellano and Bond moments are used, in which case different lags give different moments and these may match dynamics differently. Yet another alternative is Blundell and Bond's (2000) system GMM, which works with the level equation (rather than the difference equation as in Equation 21.7 above) and uses first differences of the dependent variable as instruments for the lagged level. For consistency, this estimator thus requires that the initial value of the dependent variable, in this case democracy, is uncorrelated with the fixed effects. This is unlikely to be a good assumption in our context given the historically determined nature of both democracy and inequality/redistribution.

one-step GMM estimators with a naive weighting matrix that assumes the original residuals are i.i.d.[18] Despite the potential loss in efficiency, these estimators have the advantage of being consistent when T (the time dimension of the panel) and N (the number of countries) are large, even if the number of moments also becomes large (see Alvarez and Arellano, 2003).

As the above description indicates, the source of bias in the estimation of Equation (21.6) with OLS is that the persistence parameter ρ is not estimated consistently when the time dimension does not go to infinity, and this bias translates into a bias in all other coefficient estimates. If we knew the exact value of ρ and could impose it, the rest of the parameters could be estimated consistently by OLS. Motivated by this observation, we also report OLS estimates of Equation (21.6) imposing a range of values of ρ, which shows that our main results are robust to any value of ρ between 0 and 1, increasing our confidence in the GMM estimates.

In all cases, we first focus on results using a 5-year panel, where we take an observation every 5 years from 1960 to 2010. This is preferable to taking averages, which would introduce a complex pattern of serial correlation, making consistent estimation more difficult. The 5-year panel is a useful starting point since we expect many of the results of democracy on the tax to GDP ratio (henceforth, short for tax revenue as a percentage of GDP) and inequality not to appear instantaneously or not even in one or two years. In the case of inequality measures, this is also the highest frequency we can use.[19] For the tax to GDP ratio, the annual data are available, and we also estimate annual panels, which are similar to Equation (21.6) except that in that case we include up to 12 annual lags of both the lagged dependent variable and the democracy measure on the right-hand side.

Finally, it is worth reiterating that in all of our estimates, if democracy is correlated with other changes affecting taxes or inequality, our estimates will be biased. The point of the GMM estimator is to remove the mechanical bias resulting from the presence of fixed effects and lagged dependent variables, not to estimate "causal effects." This would necessitate a credible source of variation in changes in democracy, which we do not use in this paper.

21.4.2 Data and Descriptive Statistics

We construct a yearly and a 5-year panel of 184 countries from independence or 1960, whichever is later, through to 2010, though not all variables are available for all countries

[18] When we take first differences of the data, the weighting matrix has 1 on the main diagonal and -0.5 on the subdiagonals below and above it. When we take forward orthogonal differences, the weighting matrix is the identity matrix.

[19] Our inequality data from SWIID provides yearly observations for the GINI coefficient, but they are 5-year moving averages of observations around that specific year, making them inappropriate for an annual panel.

in all periods. We extend the recent work by Papaioannou and Siourounis (2008) by constructing a new measure of democracy which combines information from Freedom House and Polity IV—two of the more widely used sources of data about political rights and democracy. We create a dichotomous measure of democracy in country c at time t, d_{ct}, as follows. First, we code a country as democratic during a given year if Freedom House codes it as "Free" or "Partially Free," and it receives a positive Polity IV score. If we only have information from one of Polity or Freedom House, we use additional information from Cheibub et al. (2010, henceforth CGV) and Boix et al. (2012, henceforth BMR). In these cases, we code an observation as democratic if either Polity is greater than 0, or Freedom House codes it as "Partially Free" or "Free" *and* at least one of CGV or BMR code it as democratic. We are interested in substantive changes in political power, and so we give priority to the expert codings of Polity and Freedom House, rather than the procedural codings of CGV and BMR.

We omit periods where a country was not independent. Finally, many of the democratic transitions captured by this algorithm are studied in detail by Papaioannou and Siourounis (2008), who code the exact date of the democratization. When we detect a democratization that is also in their sample (in the same country and generally within 4 years of the year obtained by the previous procedure), we modify our democracy dummy to match the date to which they trace back the event using historical sources.

The Papaioannou and Siourounis measure of democracy captures *permanent* changes in political institutions, and they find that this correlates with subsequent economic growth. One limitation of their measure is that they define permanent changes by looking at democratizations that are not reversed in the future, which raises the possibility of endogeneity of the definition of democracy to subsequent growth or other outcomes that stabilize democracy. In addition, it means that they have no variation coming from transitions from democracy to autocracy. Our measure retains the focus on large changes in political regimes while not using any potentially endogenous outcome to classify democratizations.

Our resulting democracy measure is a dichotomous variable capturing large changes in political institutions. Our sample contains countries that are always democratic ($d_{ct}=1$ for all years) like the United States and most OECD countries; countries that are always autocratic ($d_{ct}=0$ for all years) like Afghanistan, Angola, and China; countries that transition once and permanently into democracy like Dominican Republic in 1978, Spain in 1978, and many ex-Soviet countries after 1991. But different from Papaioannou and Siourounis, we also have countries that transition in and out of democracy such as Argentina, which is coded as democratic from 1973 to 1975, falls back to nondemocracy and then democratizes permanently in 1983. For more details on our construction of the democracy measure, see Acemoglu et al. (2013a). In Appendix B, we show robustness of our main results to other measures of democracy constructed by Cheibub et al. (2010) and Boix et al. (2012).

We combine this measure of democratization with national income statistics from the World Bank economic indicators. We use government taxes to GDP and revenues to GDP ratios measures obtained from Cullen Hendrix covering more than 127 countries yearly from 1960 to 2005 (Hendrix, 2010). These data come from a project now updated by Arbetman-Rabinowitz et al. (2011), and puts together in a consistent way information from the World Bank (for 1960–1972), the IMF Government Financial Statistics historical series, the IMF new GFS, and complementary national sources.[20] Other dependent variables we explored include secondary-schooling enrollment, agricultural shares of employment, and GDP from the World Bank; and our inequality data that will be described below.[21]

Our additional covariates include a measure of average intensity of foreign wars over the last 5 years, constructed from Polity IV and ranging from 0 (no episodes) to 10 (most intense episodes); a measure of social unrest from the SPEED project at the University of Illinois averaging the number of events over the last 5 years;[22] and the fraction of the population with at least secondary schooling from the Barro-Lee dataset. In order to explore interactions we use data on the nonagricultural share of employment in 1968 from Vanhanen (2013).[23] We also use the top 10% share of income in the United States from the World Top Incomes Database (Alvaredo et al., 2010).[24] Finally, we construct the average ratio between the share of income held by the top 10% relative to the bottom 50%, and the ratio between the share of income held by the bottom 10 relative to the bottom 50% before 2000 using the World Inequality Indicators Database. From now on we will refer to these measures as the top and bottom shares of income.[25]

There is some debate on the construction and standardization of inequality measures, particularly Gini coefficients, across countries. We use the data in the Standardized World Inequality Indicators Database (SWIID), constructed by Frederick Solt (Solt, 2009). This database uses the Luxembourg Income Study together with the World Inequality Indicators Database in order to construct a comprehensive cross-national panel of Gini coefficients that are standardized across sources and measures. One advantage of this dataset is that it provides both the net Gini, after taxes and transfers, and the gross Gini coefficients. Measuring country-level inequality is very data-demanding, and so no inequality

[20] http://thedata.harvard.edu/dvn/dv/rpc/faces/study/StudyPage.xhtml?globalId=hdl:1902.1/16845.

[21] In the Appendix A we consider manufacturing wages, compiled by Martin Rama from UNIDO statistics and averaged over 5-year intervals.

[22] http://www.clinecenter.illinois.edu/research/speed-data.html.

[23] http://www.fsd.uta.fi/en/data/catalogue/FSD1216/meF1216e.html.

[24] http://topincomes.g-mond.parisschoolofeconomics.eu/.

[25] The World Inequality Indicators Database reports income shares created using different proxies for income, including consumption, monetary income, disposable income, and others. We standardized these ratios by regressing them on a full set of dummies for each income concept and using the residuals. The raw ratios are presented only in the summary statistics.

database is completely satisfactory, but we believe the SWIID provides the most comprehensive and consistent measure for the panel regressions we are estimating. We have experimented with a number of other measures of Gini coefficients, but none have the standardized sample coverage of the SWIID. In particular, we also created a panel with data every 5 years using observations for the Gini coefficient from the World Income Inequality Database (WIID) and CEDLAS (for Latin American countries), and obtained very similar results.

Descriptive statistics for all variables used in the main sample are presented in Table 21.1, separately by our measure of nondemocracy and democracy (observations in a country that was nondemocratic at the time or democratic). In each case, we report means, standard deviations, and also the total number of observations (note that our

Table 21.1 Summary statistics

	Nondemocracies			Democracies		
Variable	**Mean**	**Std. Dev.**	***N***	**Mean**	**Std. Dev.**	***N***
Tax revenue as a percentage of GDP	15.82	9.50	660	20.94	9.73	569
Total government revenue as a percentage of GDP	20.74	12.85	660	25.42	11.01	569
Gini coefficient, net income	38.91	10.76	338	36.81	10.19	497
Gini coefficient, gross income	43.92	11.72	338	45.11	7.71	497
Foreign wars (polity)	0.15	0.70	740	0.07	0.39	623
Social unrest (SPEED)	5.35	24.99	927	9.16	35.40	705
Share with secondary enrollmenty (Barro-Lee)	17.59	16.00	745	32.07	19.23	652
Nonagricultural share of population	64.54	28.51	138	81.39	19.55	301
Nonagricultural share of GDP	74.05	16.65	627	86.32	13.47	649
Secondary enrollment	45.95	31.50	492	76.01	29.90	545
Land Gini	59.96	15.21	214	62.96	16.23	399
Nonagricultural share of population in 1968	35.60	20.94	803	56.55	25.30	598
United States top 10% income share	36.03	5.07	1050	39.43	5.47	822
Top share	1.77	1.32	81	1.34	1.06	237
Bottom share	0.10	0.03	81	0.10	0.03	237
GDP per capita in 2000 dollars	2061.78	3838.08	718	8160.03	9415.89	770

Note: Summary statistics broken by observations during nondemocracy (left panel) and democracy (right panel). See the text for a full description of the data.

sample is not balanced). The summary statistics show that democracies tend to be significantly more economically developed than nondemocracies, with much higher GDP per capita, more education, and smaller agricultural shares of employment (both on average in the sample and in 1968) and GDP. These patterns are relatively well known and are sometimes interpreted as support for modernization theory (but see Acemoglu et al., 2008, 2009 on why this cross-sectional comparison is misleading).

The differences in tax to GDP ratios and revenue to GDP ratios are much smaller; both variables are roughly 4 percentage points higher in democracies than nondemocracies, although not significantly so.[26] Consistent with this tax difference reflecting increased redistribution, after-tax inequality, measured by the net Gini, is almost three points lower in democracies, whereas pretax inequality is one point higher (the Gini is measured on a 0- to 100-scale). Figure 21.1 shows the evolution of average democracy in our sample between 1960 and 2010.[27]

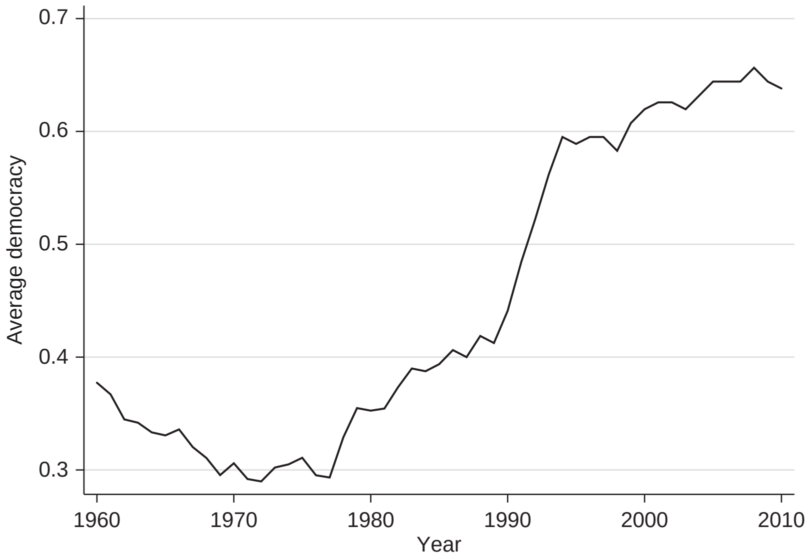


Figure 21.1 Worldwide average democracy since 1960.

[26] This comparison is broadly consistent with the cross-national regressions of Gil et al. (2004), though it is interesting that even in this cross section we do see some differences between democracies and nondemocracies.

[27] Note that democracies appear to be associated with a higher income share of the top 10% in the United States. This is because of the trend shown in Figure 21.1, making democracies more common in the recent past when this variable has also been higher.

21.5. MAIN RESULTS

21.5.1 The Effect of Democracy on Taxes

Our first results are contained in Table 21.2, which reports estimates of Equation (21.6) with the log of tax revenue to GDP ratio (tax to GDP ratio for short) as the dependent variable.

Column 1 is estimated by OLS imposing $\rho = 0$ in Equation (21.6). Though biased when $\rho > 0$, this is a natural benchmark, particularly since it corresponds to a specification often used in the literature. In all columns, we report standard errors corrected for arbitrary heteroskedasticity and serial correlation at the country level. We multiply the coefficient on democracy by 100 to ease interpretation. Throughout, we always report the number of observations, number of countries in the sample, and the number of switches in democracy from 0 to 1 or vice versa in the estimation sample (which is 92 in this case). All models include a lag of GDP per capita as a control, but the coefficients are not reported to save space. The coefficient on the estimated effect of democracy in this column, 15.00 (to two decimal places), implies a 15% increase in the tax to GDP ratio with a standard error of 4.33, and is thus statistically significant at less than the 1% confidence level. This estimate is also economically significant. It indicates that democratization—that is, a change in our democracy dummy—is associated with a 2.4 percentage points increase in the tax to GDP ratio.

Column 2 includes the lag of tax to GDP ratio on the right-hand side, thus relaxing the assumption that $\rho = 0$. The effect of democracy, γ, is now estimated to be 11.7 (approximately 11.7%, with standard error = 3.38) and is again statistically significant at less than the 1% level. In the presence of the lagged dependent variable on the right-hand side in this specification, γ is now merely the short-run impact of democracy on the tax to GDP ratio, not the long-run effect. The estimate of ρ is 0.27, and is significant, suggesting that there is indeed some persistence in the dependent variable. To obtain the long-run effect, we set $z_{it} = z_{it-1}$ so that the dynamics in the outcome variable converge to the new "steady state." This gives the long-run effects of a switch to democracy as

$$\frac{\gamma}{1-\rho},$$

and is reported at the bottom, together with the p-value for the hypothesis that it is equal to 0. In Column 2, this long-run effect implies a 16% increase in the tax to GDP ratio from a permanent switch to democracy.

Figure 21.2 shows the effect of democracy on the tax to GDP ratio visually. Here, similar to an event study analysis, we place all transitions to democracy at $t = 0$, and those observations before then (with $t < 0$) show the trends in tax to GDP ratio before democratization, and those with $t > 0$ correspond to changes in the tax to GDP ratio after

Table 21.2 Effects of democratization on the log of tax revenue as a percentage of GDP

			GMM			Assuming AR(1) coefficient				
						$\rho=0$	$\rho=0.25$	$\rho=0.5$	$\rho=0.75$	$\rho=1$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Democracy lagged	15.00⋆⋆⋆	11.71⋆⋆⋆	11.27	18.68⋆⋆	14.63⋆⋆	15.00⋆⋆⋆	11.92⋆⋆⋆	8.84⋆⋆⋆	5.77⋆⋆	2.69
	(4.33)	(3.38)	(7.23)	(8.78)	(5.98)	(4.33)	(3.27)	(2.55)	(2.48)	(3.11)
Dep. Var. lagged		0.27⋆⋆⋆	0.27⋆⋆⋆	0.29⋆⋆⋆	0.33⋆⋆⋆					
		(0.06)	(0.10)	(0.07)	(0.08)					
Observations	944	944	816	816	816	944	944	944	944	944
Countries	128	128	125	125	125	128	128	128	128	128
Number of moments			81	61	61					
Hansen p-value			0.12	0.05	0.06					
AR2 p-value			0.92	0.83	0.78					
Democracy changes in the sample	92	92	82	82	82	92	92	92	92	92
Long-run effect of democracy	15.00	15.97	15.49	26.35	21.97	15.00	15.89	17.68	23.06	.
p-Value for the long-run effect	0.00	0.00	0.11	0.03	0.01	0.00	0.00	0.00	0.02	.

Note: OLS estimates (Columns 1–2) include a full set of country and year fixed effects. Arellano and Bond's GMM estimators of the dynamic panel model (Columns 3–4) remove country fixed effects by taking first differences of the data, or by taking forward orthogonal differences (Column 5) and then constructing moment conditions using predetermined lags of the dependent variable and democracy. Columns 4 and 5 use up to the fifth lag of predetermined variables to create moments, restricting the number of moments used. Columns 6–10 impose different values for the autocorrelation coefficient in the tax revenue as a percentage of GDP series, and estimates the effect of democracy including a full set of country and year fixed effects. All models control for lagged GDP per capita but this coefficient is not reported to save space. Robust standard errors, adjusted for clustering at the country level, are in parentheses. ⋆⋆⋆: significant at 1%; ⋆⋆: significant at 5%; ⋆: significant at 10%. We do not report long-run effects and their p-values in Column 10 because they are not defined for $\rho=1$.

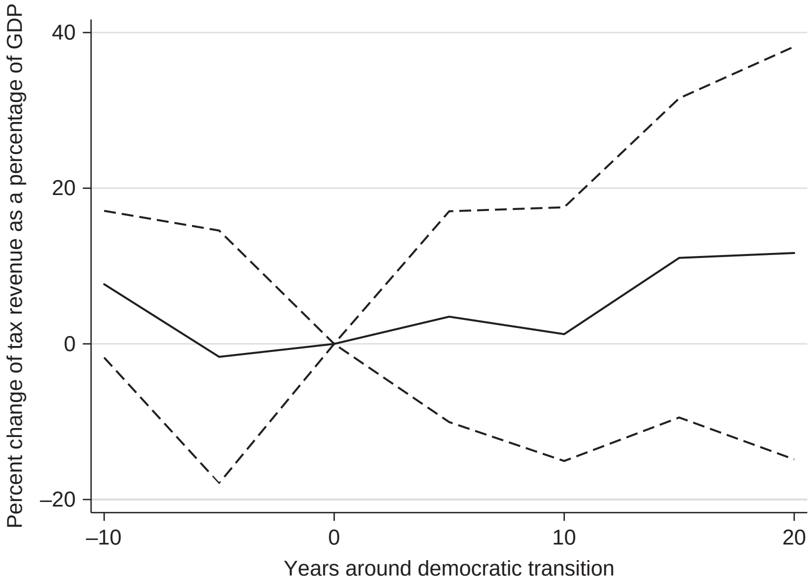


Figure 21.2 Tax revenue as a percentage of GDP around a democratization. Constructed using the 5-year panel.

democratization. The figure shows that there is no discernible change in the tax to GDP ratio before democratization, increasing our confidence in the results concerning the effect of democracy on taxes. It also confirms that the effect of democracy on the tax to GDP ratio evolves only slowly, reaching a maximum 15 years after the democratization takes place. This underscores the role of the lagged dependent variable in our econometric specifications.

As a second diagnostic for our estimates, Figure 21.3 shows a scatterplot of the residuals of the tax to GDP ratio (in logs) on the vertical axis against the residuals of the lag of our democracy measure on the horizontal axis. All covariates, including year and country fixed effects, and the lagged dependent variable, are partialed out. Each point corresponds to a particular country/year observation. The slope of the regression line coincides with our estimated coefficient of 11.7. The figure shows that the estimated relationship does not seem to be driven by any particular outlier. To explore this more formally we removed 49 observations whose Cook distance was above the rule of thumb $4/N$, with N the sample size and reestimated our model. The coefficient of democracy falls to 8.28 with standard error 2.46, and is still significant at the 1% level. The bottom panel of Figure 21.3 shows the scatterplot excluding these outliers. We have experimented with a number of other methods for dealing with outliers, such as Huber M-regressions and excluding outliers with estimated standardized errors >1.96, and our results on tax to GDP ratios remain generally unchanged.

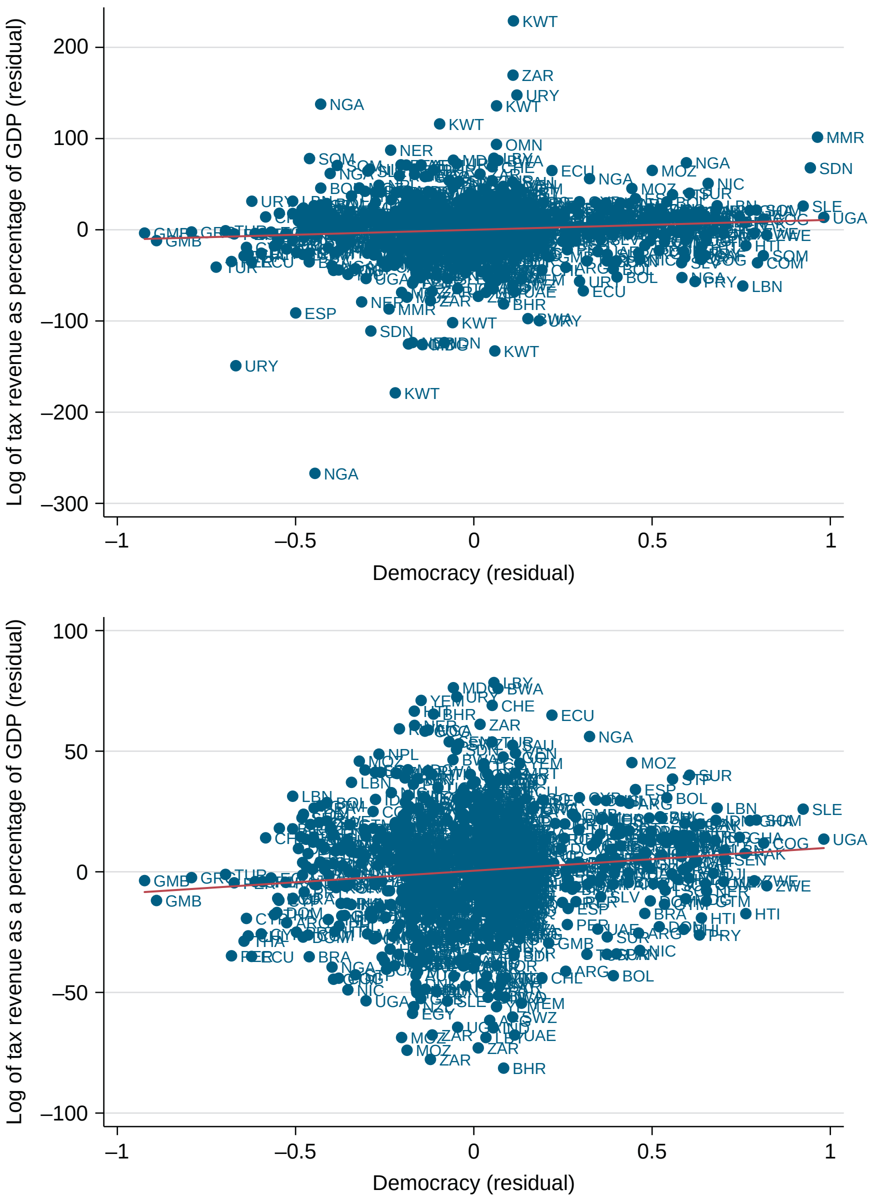


Figure 21.3 Residual of tax revenue as a percentage of GDP (vertical axis) against the residual of our democracy indicator. Each dot is a country/year observation, and there are a total of 975 observations. Bottom figure excludes outliers.

As noted in the previous subsection, the OLS estimator of Column 2 is inconsistent because of the (downward) bias in the estimation of ρ. Column 3 reports the GMM estimator described earlier with the full set of moments (in this case, this corresponds to 82 moments as noted in the table). Notably, the estimate for ρ is identical up to two decimal places, indicating in fact that if there was a downward bias in the estimation of Column 2, it was negligible, suggesting that the large-T assumption (given the low persistence ρ) is a good approximation. The estimate for γ also decreases marginally, but the standard error increases substantially, making the resulting estimate insignificant at conventional levels. However, the long-run impact is very similar to the OLS estimate of approximately 15 (15%), with a p-value of 0.11. It should also be noted that the tests for second-order autocorrelation in the error term and the Hansen's J test for over identification pass comfortably, thus further increasing our confidence in this specification.

Columns 4 and 5 present alternative GMM estimators with fewer moments and with forward-differencing, respectively. Both estimates only use up to the fifth valid lags of democracy and the dependent variable to form moment conditions. The point estimates on both γ and ρ are larger than Columns 2 and 3, and significant at the 5% level, and hence imply the significantly larger long-run effects, 26% and 21%, respectively, reported at the bottom.

Columns 6–10 estimate Equation (21.6), imposing different values for ρ spanning the entire interval from 0 to 1. We use the same sample as in Column 2, which is also the same one as in Column 1 and thus implies that in this case Column 6, which sets $\rho = 0$, is identical to Column 1 (this will not be the case in some of our later tables). As noted above, the problem with the OLS estimation (with fixed effects) stems from the bias in the estimate of ρ, so conditional on the correct value for this variable, the OLS estimate of the impact of democracy is consistent. In almost all cases, with the exception of the last column, there is a statistically and economically significant impact of democracy on the tax to GDP ratio. The long-run impact is smaller when ρ is assumed to take a small value, and comparable to that in Column 2 when we impose $\rho = 0.25$. The coefficient gets smaller and less significant the farther the imposed value of ρ is from the estimated values in Columns 2–5.[28] In sum, the median estimated long-run effect of democracy on the tax to GDP ratio from this table is almost 16%, with estimates that range from 15% to 26%.

Table 21.3 has the same structure as Table 21.2, but uses total government revenue to GDP ratio as the dependent variable. Though the impact of democracy is a little smaller, the pattern is qualitatively very similar, with slightly larger long-run effects in the GMM estimators relative to the OLS estimators. The estimates in Column 2 show that the coefficient of lagged democracy is 7.55 (standard error = 2.35), which is significant at the

[28] In Column 10 where we impose $\rho = 1$, we do not compute the long-run impact, since this is undefined in this unit-root specification. The coefficient in this specification is small and insignificant, suggesting that there is not much variation in growth rates of tax to GDP to be explained by democratization.

Table 21.3 Effects of democratization on the log of total government revenue as a percentage of GDP

			GMM			Assuming AR(1) coefficient				
						$\rho=0$	$\rho=0.25$	$\rho=0.5$	$\rho=0.75$	$\rho=1$
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
Democracy lagged	9.31***	7.55***	9.37*	11.13**	10.04**	9.31***	8.06***	6.81***	5.56***	4.31
	(3.44)	(2.35)	(5.01)	(5.58)	(4.37)	(3.44)	(2.60)	(2.08)	(2.15)	(2.76)
Dep. Var. lagged		0.35***	0.47***	0.52***	0.53***					
		(0.03)	(0.06)	(0.06)	(0.06)					
Observations	944	944	816	816	816	944	944	944	944	944
Countries	128	128	125	125	125	128	128	128	128	128
Number of moments			81	61	61					
Hansen p-value			0.05	0.04	0.05					
AR2 p-value			0.36	0.39	0.40					
Democracy changes in the sample	92	92	82	82	82	92	92	92	92	92
Long-run effect of democracy	9.31	11.64	17.77	22.96	21.47	9.31	10.74	13.61	22.23	.
p-Value for the long-run effect	0.01	0.00	0.07	0.05	0.03	0.01	0.00	0.00	0.01	.

Note: OLS estimates (Columns 1–2) include a full set of country and year fixed effects. Arellano and Bond's GMM estimators of the dynamic panel model (Columns 3–4) remove country fixed effects by taking first differences of the data, or by taking forward orthogonal differences (Column 5) and then constructing moment conditions using predetermined lags of the dependent variable and democracy. Columns 4 and 5 use up to the fifth lag of predetermined variables to create moments, restricting the number of moments used. Columns 6–10 impose different values for the autocorrelation coefficient in the total government revenue as a percentage of GDP series, and estimates the effect of democracy including a full set of country and year fixed effects. All models control for lagged GDP per capita but this coefficient is not reported to save space. Robust standard errors, adjusted for clustering at the country level, are in parentheses. ***: significant at 1%; **: significant at 5%; *: significant at 10%. We do not report long-run effects and their p-values in Column 10 because they are not defined for $\rho=1$.

1% level. The long-run effect of democracy is to increase total revenue as a percentage of GDP by 11.64 and is significant at the 1% level. The baseline GMM estimator leads to larger values of ρ and γ, resulting in a larger long-run effect of 17.8%. Figure 21.4 is the analogue of Figure 21.2, but using the total revenue to GDP ratio measure instead, and shows a similar pattern, although there is a slight downward trend prior to democracy in this variable. In sum, the evidence again suggests that democracy results in larger government revenues as a share of GDP.

Table 21.4 estimates Equation (21.6) for the annual panel. Column 1 includes just four (annual) lags of the dependent variable and democracy on the right-hand side, and is estimated by OLS. Even though individual lags of democracy are not significant, they are jointly significant as witnessed by the long-run effect reported at the bottom, which is similar to the OLS long-run effect in Table 21.2. Column 2 adds four more lags and Column 3 adds four further lags, for a total of 12 lags of democracy and the dependent variable on the right-hand side (to economize on space, we only report the p-values for F-tests for the joint significance of these additional lags). The overall pattern and the long-run effects are very similar to Column 1. Columns 4–6 estimate the same models using the Arellano and Bond GMM estimator. The long-run effects are substantially higher and comparable to the one estimated in Columns 3–5 in Table 21.2 using the 5-year panel.

Table 21.5 probes the robustness of the tax to GDP ratio results, focusing on the 5-year panel. Odd-numbered columns report OLS estimates of Equation (21.6), whereas even-numbered columns are for the GMM estimator (equivalent to Column 3 of

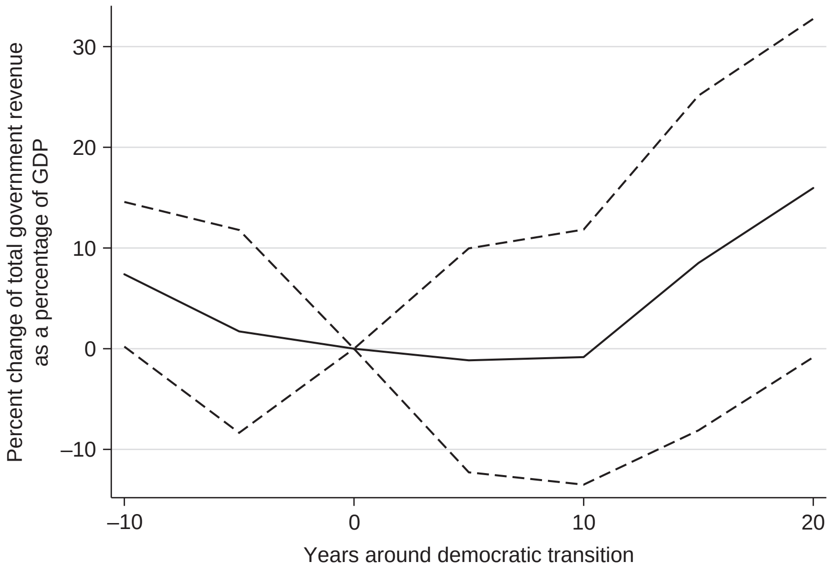


Figure 21.4 Total government revenue as a percentage of GDP around a democratization. Constructed using the 5-year panel.

Table 21.4 Effects of democratization on the log of tax revenue as a percentage of GDP, yearly panel

	OLS			GMM		
	(1)	(2)	(3)	(4)	(5)	(6)
D_{t-1}	3.43	3.45	4.06	5.49	9.49	8.11⋆
	(2.82)	(2.91)	(3.13)	(3.83)	(5.82)	(4.86)
D_{t-2}	−2.31	−2.01	−2.08	−1.66	−0.67	−1.04
	(2.83)	(2.86)	(3.00)	(2.67)	(2.56)	(2.86)
D_{t-3}	0.66	−0.03	−1.66	1.25	0.53	−0.88
	(2.21)	(2.36)	(2.58)	(2.24)	(2.21)	(2.46)
D_{t-4}	1.65	2.83	3.88⋆	6.14⋆⋆⋆	3.45⋆	4.29⋆⋆
	(1.63)	(2.03)	(2.14)	(2.32)	(1.81)	(1.93)
p-Value, first four democracy lags	0.02	0.10	0.20	0.06	0.07	0.06
p-Value, second four democracy lags	.	0.61	0.21	.	0.13	0.09
p-Value, third four democracy lags	.	.	0.80	.	.	0.82
y_{t-1}	0.65⋆⋆⋆	0.64⋆⋆⋆	0.62⋆⋆⋆	0.58⋆⋆⋆	0.53⋆⋆⋆	0.52⋆⋆⋆
	(0.04)	(0.04)	(0.05)	(0.05)	(0.04)	(0.06)
y_{t-2}	0.06	0.08	0.09⋆	0.03	0.04	0.06
	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.05)
y_{t-3}	0.09	0.10	0.12	0.07	0.07	0.11
	(0.09)	(0.09)	(0.10)	(0.08)	(0.09)	(0.09)
y_{t-4}	−0.00	−0.03	−0.06	−0.03	−0.03	−0.06
	(0.06)	(0.06)	(0.06)	(0.07)	(0.05)	(0.05)
p-Value, first four tax to GDP lags	0.00	0.00	0.00	0.00	0.00	0.00
p-Value, second four tax to GDP lags	.	0.61	0.55	.	0.05	0.19
p-Value, third four tax to GDP lags	.	.	0.51	.	.	0.11
Observations	4434	3925	3425	4306	3799	3301
Countries	128	126	124	128	125	123
Number of moments				373	637	837
Hansen p-value				1.00	1.00	1.00
AR2 p-value				0.30	0.39	0.96
Democracy changes in sample	75	73	69	75	73	68
Long-run effect of democracy	16.49	19.11	12.49	32.38	38.85	25.40
p-Value for long-run effect	0.00	0.00	0.06	0.03	0.01	0.02

Note: OLS estimates (Columns 1–3) include a full set of country and year fixed effects. Arellano and Bond's GMM estimators of the dynamic panel model (Columns 4–6) remove country fixed effects by taking first differences of the data and then constructing moment conditions using as many predetermined lags of the dependent variable and democracy as included in the model. To save space we only report the p-value of a joint test of significance for lags 5–8 (second four lags) and lags 9–12 (third four lags). All models control for as many lags of GDP per capita as lags of democracy in the equation, but these coefficients are not reported to save space. ⋆⋆⋆: significant at 1%; ⋆⋆: significant at 5%; ⋆: significant at 10%. We do not report long-run effects and their p-values in Column 10 because they are not defined for $\rho = 1$.

Table 21.5 Effects of democratization on the log of tax revenue as a percentage of GDP with different set of controls

	Ex. GDP per capita		Adding other controls							
	OLS	GMM	OLS	GMM	OLS	GMM	OLS	GMM	OLS	GMM
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Democracy lagged	10.91★★★	12.59★	12.22★★★	12.42★	11.70★★★	10.73	11.59★★	15.01★★	11.68★★★	15.34★★
	(3.69)	(6.67)	(3.52)	(6.73)	(3.38)	(7.00)	(3.46)	(7.59)	(3.48)	(6.90)
Dep. Var. lagged	0.28★★★	0.31★★★	0.27★★★	0.28★★★	0.27★★★	0.28★★★	0.31★★★	0.32★★★	0.31★★★	0.34★★★
	(0.07)	(0.11)	(0.06)	(0.10)	(0.06)	(0.10)	(0.07)	(0.12)	(0.07)	(0.12)
War lagged			−1.60	−2.38					−2.30	−4.26
			(2.56)	(3.91)					(3.03)	(4.06)
Unrest lagged					0.01	−0.06			0.01	−0.09
					(0.02)	(0.07)			(0.02)	(0.08)
Education lagged							−0.16	0.02	−0.15	−0.30
							(0.19)	(0.63)	(0.20)	(0.69)
Observations	1090	957	889	771	927	802	844	734	803	700
Countries	133	133	118	115	125	122	110	107	103	100
Number of moments		80		82		82		82		84
Hansen p-value		0.22		0.07		0.17		0.15		0.21
AR2 p-value		0.24		0.88		0.91		0.76		0.77
Democracy changes in the sample	101	90	89	80	92	82	77	68	77	68
Long-run effect of democracy	15.22	18.26	16.64	17.27	15.97	14.84	16.76	22.15	16.97	23.17
p-Value for the long-run effect	0.00	0.07	0.00	0.06	0.00	0.12	0.00	0.04	0.00	0.02

Note: OLS estimates (odd columns) include a full set of country and year fixed effects. Columns 3–10 include lagged GDP per capita as a control. Arellano and Bond's GMM estimators of the dynamic panel model (even columns) remove country fixed effects by taking first differences of the data and then constructing moment conditions using predetermined lags of the dependent variable and democracy. Robust standard errors, adjusted for clustering at the country level, are in parentheses. ★★★★: significant at 1%; ★★: significant at 5%; ★: significant at 10%.

Table 21.2). The first two columns exclude GDP per capita as a control. Reassuringly, however, our coefficients remain positive and significant, implying a 10–15% increase in the tax to GDP ratio following a democratization. Columns 3 and 4 include the lagged index of foreign wars. This is useful since several authors have claimed that either in history or in the recent past, war has been a major determinant of taxation and redistribution policies. For example, the famous Tilly (1985) hypothesis explains the growth of the state with war and preparation for war (see also Besley and Persson, 2011). More recently, Atkinson et al. (2011) have pointed to large wars and the concomitant economic changes as some of the most significant events correlated with declines of 1% income shares in combatant countries (see also Scheve and Stasavage, 2010, 2012). In contrast to these hypotheses, we do not find any effect of war on the tax to GDP ratio in our post-war panel. The effect of democracy on the tax to GDP ratio remains essentially unchanged when the external war index is included.

Columns 5 and 6 include the lagged measure of social unrest from the SPEED data. This variable is insignificant and has no effect on the coefficient of democracy. Columns 7 and 8 include the stock of education, measured as the fraction of the population with at least secondary schooling from the Barro-Lee dataset, which could be an important determinant of fiscal policy and inequality. Once again, this variable has no major effect on the estimate of the impact of democracy on the tax to GDP ratio and is itself insignificant. Columns 9 and 10 include all three of these variables together, again with a very limited impact on our estimates and no evidence of an effect on war, unrest or the stock of education. The long-run effects at the bottom are very similar to those in Table 21.2 and highly significant.[29]

Overall, the evidence in Tables 21.2–21.5 shows a strong and robust impact of democracy on taxes as measured by the tax to GDP ratio or the government revenue to GDP ratio. This evidence suggests that democracy does lead to more taxes. This evidence is consistent with several of the works discussed above, though it is in stark contrast with Gil et al. (2004). The main difference is the cross-national focus of Gil, Mulligan, and Sala-i-Martin, which contrasts with our econometric approach exploiting the within-country variation (with country fixed effects and also controlling for the dynamics of the tax to GDP ratio). For reasons explained above, we believe that the cross-sectional relationship is heavily confounded by other factors and is unlikely to reveal much about the impact of democracy on redistribution and taxes.

We next investigate whether there is an impact of democracy on inequality.

[29] Another relevant robustness check is to include ex-Soviet countries in the sample. However, fiscal data are only available for Hungary, Poland, and Romania, and then only for the 1990–1995 period, which results in the observations being absorbed by the fixed effects. We thus do not report this robustness check for these specifications (but will report it for our inequality results).

21.5.2 The Effect of Democracy on Inequality

Tables 21.6 and 21.7 turn to the effect of democracy on inequality. Each panel of Table 21.6 mirrors Table 21.2, with the top panel using the net Gini coefficient (after tax and transfers) and the bottom panel using the gross Gini coefficient (before tax and redistribution) as dependent variables.

Though the sample is smaller and data quality may be lower, the most important message from these tables is that there is no consistent evidence for a significant effect of democracy on inequality. Some of our specifications show negative effects of democracy on inequality, particularly on the gross Gini coefficient, but these tend to have large standard errors and are not stable across specifications.

For example, in Table 21.6, most of our estimates suggest there is a negative effect of democracy on the net Gini coefficient, but none of these estimates is statistically significant at the standard levels. For instance, the estimates in Column 3 imply that democracy reduces the Gini coefficient (measured on a 0- to 100-scale) by 2.01 points (standard error = 1.59) in the short run, and by 3.1 points in the long run. Given the standard deviation of the net Gini of 10.76 (see Table 21.1), these effects are quantitatively sizable (though they are also smaller in other columns) but also statistically insignificant. The magnitudes for the gross Gini are similar, but a few specifications contain significant results (those with imposed values of $\rho > 0.5$). This may be because there is less measurement error in this measure relative to the net Gini, which does depend on potentially misreported taxes and transfers.

The AR2 test for the GMM estimator for the net Gini suggests there is higher order autocorrelation in the transformed errors, which invalidates the use of second lags as instruments. However, when we only use deeper lags to form valid moment conditions we get very similar results, with smaller effects of democracy on inequality, consistent with the fact that the Hansen overidentification test passes comfortably. The specification tests (AR2 and Hansen J test) for our models using the gross Gini as dependent variable also pass comfortably.

Figure 21.5, which is similar to Figures 21.2 and 21.4, visually shows that there is no substantial fall in inequality following a democratization. There is no pre-trend in inequality. But there is a temporary increase in inequality prior to democratization, which could have persistent effects biasing our estimates unless we control for the dynamics of inequality, further motivating our specifications controlling for such dynamics.

As a second diagnostic of our estimates, Figure 21.6 again shows a scatterplot of the residuals of the net Gini on the vertical axis against the residuals of the lag of our democracy measure on the horizontal axis. All covariates, including year and country fixed effects and the lagged dependent variable, are partialed out. Each point corresponds to a particular country/year observation. The slope of the regression line coincides with our estimated coefficient of −0.744 in Column 2 of the top panel in Table 21.6. The

Table 21.6 Effects of democratization on inequality

			GMM			Assuming AR(1) coefficient				
						$\rho=0$	$\rho=0.25$	$\rho=0.5$	$\rho=0.75$	$\rho=1$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent variable: Gini coefficient, net income										
Democracy lagged	0.62	−0.74	−2.01	−2.60	−1.60	−0.42	−0.67	−0.92	−1.17	−1.42
	(0.78)	(0.88)	(1.59)	(1.63)	(1.51)	(0.93)	(0.89)	(0.89)	(0.93)	(1.00)
Dep. Var. lagged		0.32***	0.35***	0.39***	0.32***					
		(0.07)	(0.10)	(0.12)	(0.12)					
Observations	657	537	420	420	424	537	537	537	537	537
Countries	127	113	100	100	100	113	113	113	113	113
Number of moments			81	61	61					
Hansen p-value			0.60	0.69	0.30					
AR2 p-value			0.02	0.03	0.01					
Democracy changes	65	47	31	31	31	47	47	47	47	47
Long-run effect	0.62	−1.10	−3.12	−4.28	−2.36	−0.42	−0.90	−1.84	−4.67	.
p-Value	0.43	0.40	0.21	0.12	0.30	0.65	0.45	0.31	0.21	.
Dependent variable: Gini coefficient, gross income										
Democracy lagged	−1.22	−1.50	−1.45	−1.88	−1.22	−1.51	−1.50	−1.50*	−1.49*	−1.49
	(0.99)	(0.90)	(1.44)	(1.59)	(1.27)	(1.15)	(1.00)	(0.90)	(0.87)	(0.92)
Dep. Var. lagged		0.50***	0.64***	0.64***	0.76***					
		(0.06)	(0.11)	(0.11)	(0.11)					
Observations	657	537	420	420	424	537	537	537	537	537
Countries	127	113	100	100	100	113	113	113	113	113
Number of moments			81	61	61					
Hansen p-value			0.54	0.29	0.37					
AR2 p-value			0.59	0.57	0.48					
Democracy changes	65	47	31	31	31	47	47	47	47	47
Long-run effect	−1.22	−2.98	−3.99	−5.26	−5.15	−1.51	−2.00	−3.00	−5.97	.
p-Value	0.22	0.11	0.36	0.30	0.42	0.19	0.14	0.10	0.09	.

Note: OLS estimates (Columns 1–2) include a full set of country and year fixed effects. Arellano and Bond's GMM estimators of the dynamic panel model (Columns 3–4) remove country fixed effects by taking first differences of the data, or by taking forward orthogonal differences (Column 5) and then constructing moment conditions using predetermined lags of the dependent variable and democracy. Columns 4 and 5 use up to the fifth lag of predetermined variables to create moments, restricting the number of moments used. Columns 6–10 impose different values for the autocorrelation coefficient of the dependent variable, and estimates the effect of democracy including a full set of country and year fixed effects. All models control for lagged GDP per capita but this coefficient is not reported to save space. Robust standard errors, adjusted for clustering at the country level, are in parentheses. ***: significant at 1%; **: significant at 5%; *: significant at 10%. We do not report long-run effects and their p-values in Column 10 because they are not defined for $\rho=1$.

Table 21.7 Effects of democratization on inequality adding controls

	Ex GDP per capita		Baseline sample								Inc. Ex-Soviets	
	OLS	GMM	OLS	GMM	OLS	GMM	OLS	GMM	OLS	GMM		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dependent variable: Gini coefficient, net income												
Democracy lagged	−0.87	−2.81⋆⋆	−0.71	−1.87	−0.75	−2.16	−0.72	−1.46	−0.72	−1.69	−0.26	−1.51
	(0.82)	(1.31)	(0.93)	(1.68)	(0.88)	(1.58)	(1.03)	(1.87)	(1.06)	(1.86)	(0.77)	(1.32)
Dep. Var. lagged	0.33⋆⋆⋆	0.49⋆⋆⋆	0.32⋆⋆⋆	0.34⋆⋆⋆	0.32⋆⋆⋆	0.36⋆⋆⋆	0.32⋆⋆⋆	0.34⋆⋆⋆	0.32⋆⋆⋆	0.33⋆⋆⋆	0.31⋆⋆⋆	0.53⋆⋆⋆
	(0.07)	(0.10)	(0.07)	(0.11)	(0.07)	(0.10)	(0.08)	(0.12)	(0.08)	(0.12)	(0.06)	(0.10)
War lagged			0.12						0.33	0.27		
			(0.28)						(0.28)	(0.49)		
Unrest lagged					−0.01	0.00			−0.00	0.00		
					(0.00)	(0.01)			(0.00)	(0.01)		
Education lagged							−0.02	0.06	−0.02	0.01		
							(0.04)	(0.13)	(0.05)	(0.16)		
Observations	556	435	512	402	523	409	502	399	480	382	611	473
Countries	115	103	106	95	110	97	100	91	95	87	134	121
Number of moments		80		82		82		82		84		81
Hansen p-value		0.82		0.76		0.59		0.67		0.77		0.42
AR2 p-value		0.03		0.04		0.02		0.04		0.04		0.03
Democracy changes	49	34	44	30	47	31	38	28	37	27	61	39
Long-run effect	−1.30	−5.55	−1.06	−2.85	−1.10	−3.39	−1.05	−2.19	−1.06	−2.51	−0.37	−3.23
p-Value	0.29	0.03	0.44	0.26	0.39	0.18	0.48	0.44	0.49	0.37	0.74	0.27
Dependent variable: Gini coefficient, gross income												
Democracy lagged	−1.51⋆	−2.18⋆	−1.57⋆	−1.90	−1.39	−1.39	−1.80⋆	−1.29	−1.70⋆	−1.28	−0.97	−0.79
	(0.89)	(1.24)	(0.95)	(1.52)	(0.89)	(1.40)	(1.00)	(1.65)	(1.02)	(1.65)	(0.84)	(1.41)
Dep. Var. lagged	0.53⋆⋆⋆	0.75⋆⋆⋆	0.50⋆⋆⋆	0.60⋆⋆⋆	0.49⋆⋆⋆	0.65⋆⋆⋆	0.49⋆⋆⋆	0.62⋆⋆⋆	0.49⋆⋆⋆	0.62⋆⋆⋆	0.49⋆⋆⋆	0.72⋆⋆⋆
	(0.06)	(0.09)	(0.06)	(0.12)	(0.06)	(0.11)	(0.06)	(0.12)	(0.07)	(0.12)	(0.06)	(0.08)
War lagged			0.06	−0.03					0.21	−0.03		
			(0.26)	(0.44)					(0.27)	(0.46)		

Unrest lagged					−0.01★★	−0.00			−0.01★★	−0.00		
					(0.00)	(0.01)			(0.00)	(0.01)		
Education lagged							0.02	−0.02	0.02	−0.02		
							(0.06)	(0.12)	(0.06)	(0.16)		
Observations	556	435	512	402	523	409	502	399	480	382	611	473
Countries	115	103	106	95	110	97	100	91	95	87	134	121
Number of moments		80		82		82		82		84		81
Hansen *p*-value		0.51		0.70		0.45		0.79		0.84		0.28
AR2 *p*-value		0.50		0.52		0.45		0.66		0.50		0.50
Democracy changes	49	34	44	30	47	31	38	28	37	27	61	39
Long-run effect	−3.19	−8.69	−3.15	−4.72	−2.75	−3.95	−3.56	−3.38	−3.34	−3.43	−1.91	−2.84
p-Value	0.11	0.13	0.11	0.25	0.13	0.37	0.08	0.46	0.10	0.45	0.26	0.59

Note: OLS estimates (odd columns) include a full set of country and year fixed effects. Columns 3–12 control for lagged GDP per capita. Arellano and Bond's GMM estimators of the dynamic panel model (even columns) remove country fixed effects by taking first differences of the data and then constructing moment conditions using predetermined lags of the dependent variable and democracy. Robust standard errors, adjusted for clustering at the country level, are in parentheses. ★★★: significant at 1%; ★★: significant at 5%; ★: significant at 10%.

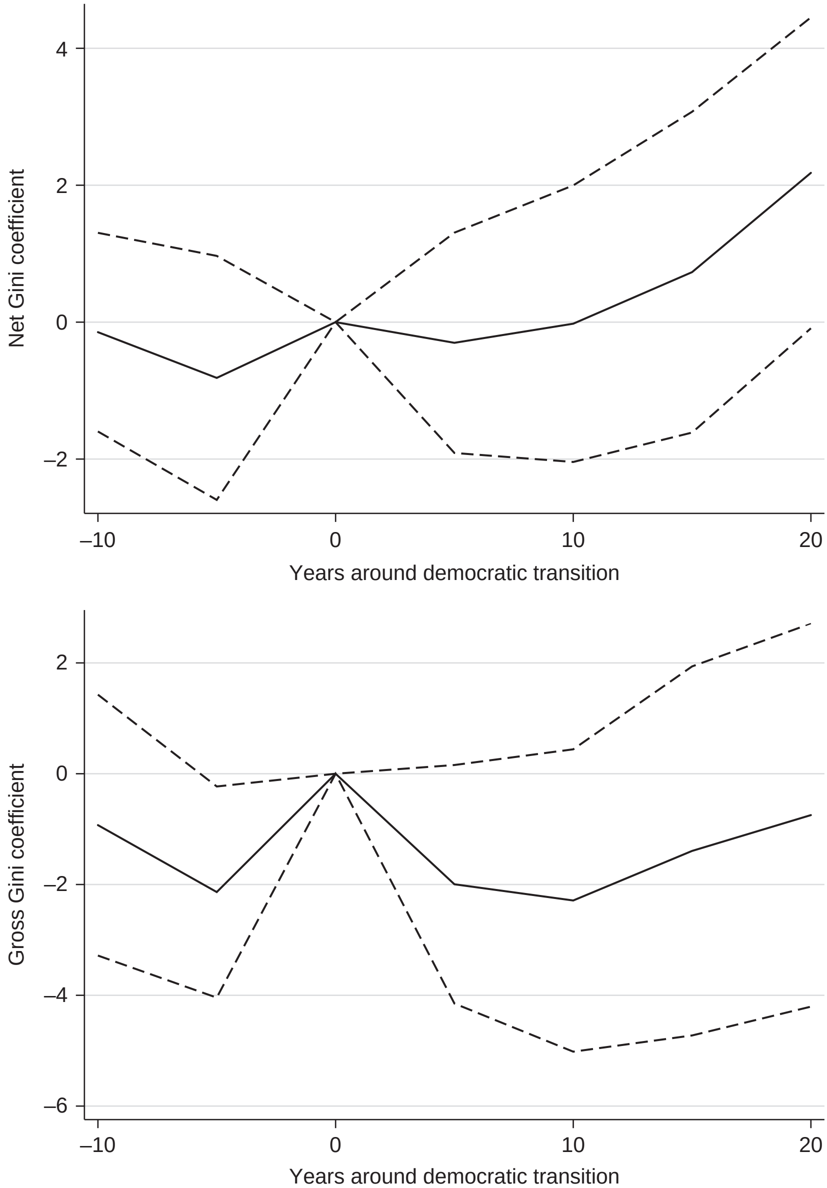


Figure 21.5 Gini coefficient around a democratization. Constructed using the 5-year panel.

figure shows that the estimated relationship does not seem to be driven by any particular outlier. Figure 21.7 shows the same scatterplot, except with gross Gini on the y-axis, and again suggests a negative, if imprecise, relationship. We explored the impact of outliers further, using a procedure similar to the one we used before. We therefore removed observations whose Cook distance was above the rule of thumb $4/N$, with N the sample

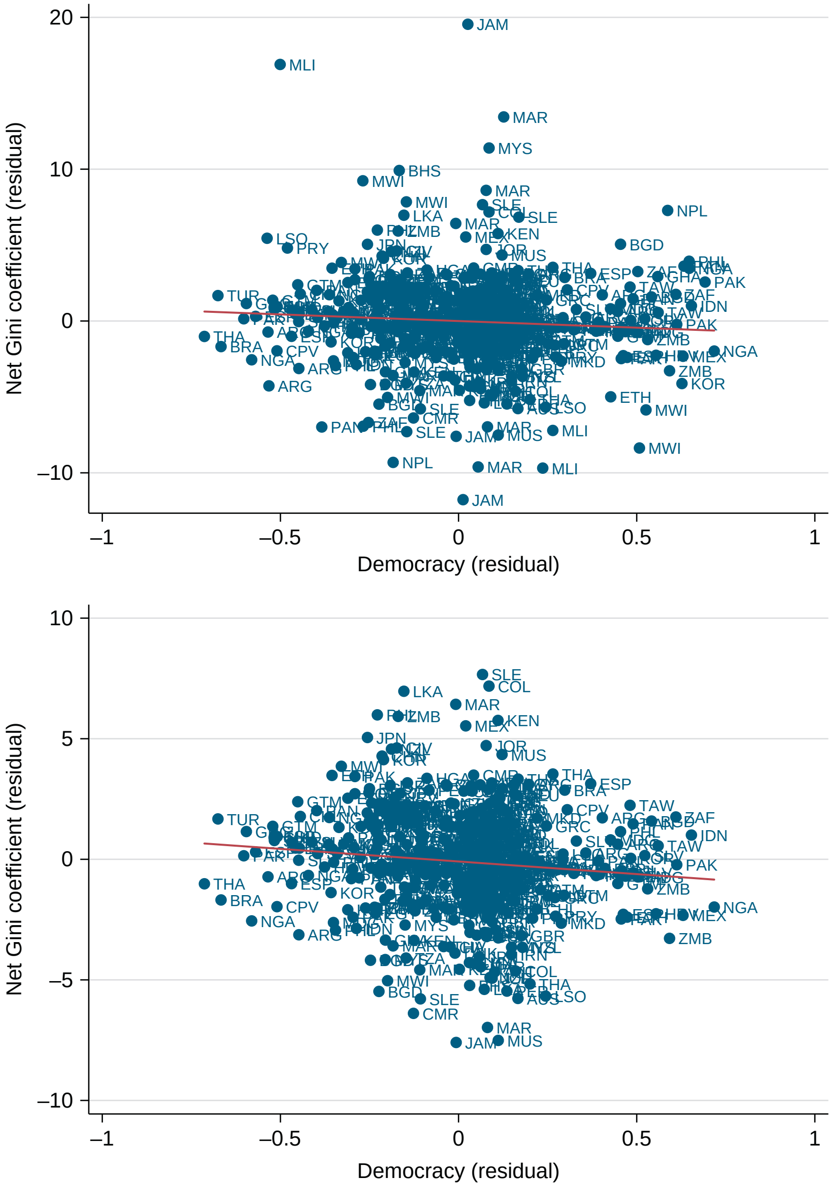


Figure 21.6 Residual of net Gini (vertical axis) against the residual of our democracy indicator. Each dot is a country/year observation, and there are a total of 538 observations. The bottom figure excludes outliers.

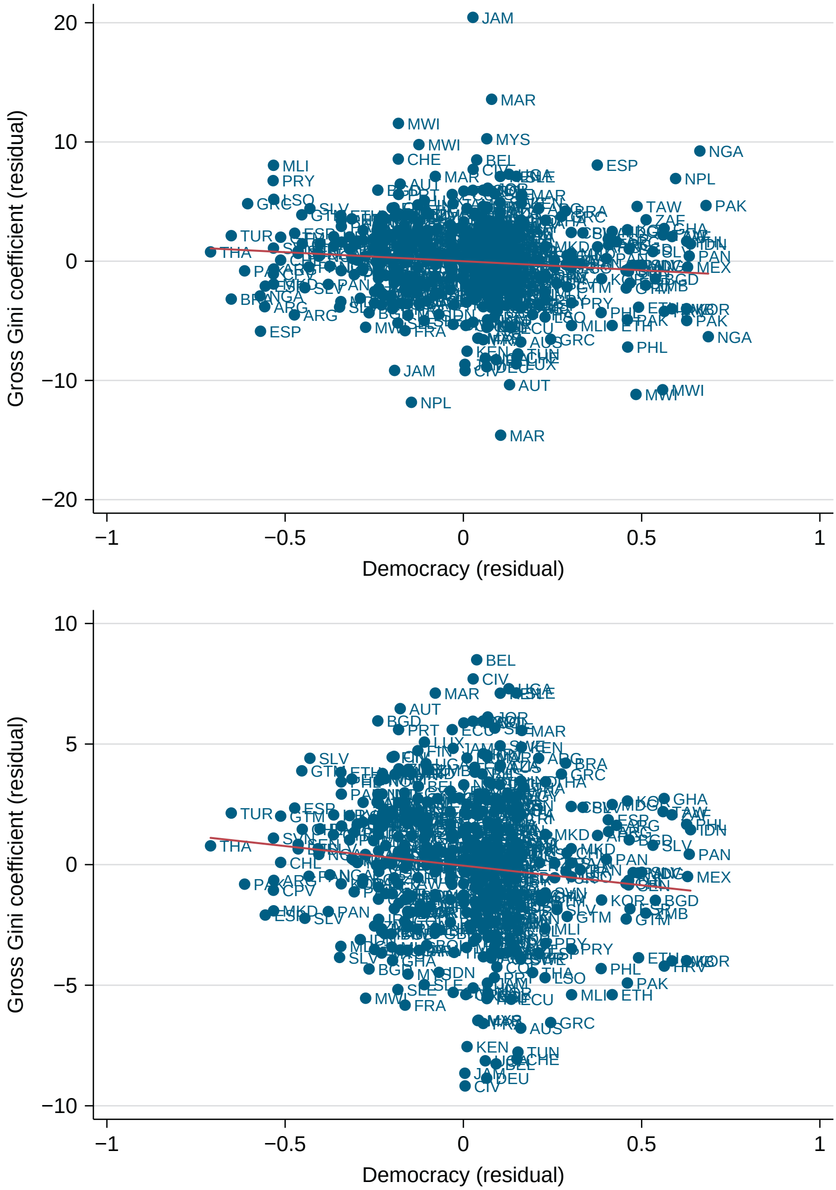


Figure 21.7 Residual of gross Gini (vertical axis) against the residual of our democracy indicator. Each dot is a country/year observation, and there are a total of 538 observations. The bottom figure excludes outliers.

size and reestimated our model. Democracy has no significant effect in this sample without the outliers for the net Gini, but there is a moderately significant effect on the gross Gini in some specifications. In addition, we found a marginally significant effect on both the net and the gross Gini when we used Huber's M estimator. When excluding observations with standardized residuals >1.96, we again found a significant negative effect on the gross Gini but not on the net Gini.

Table 21.7 adds covariates, as in Table 21.5, for the tax variables, and comprises two panels, one for each Gini measure. The only difference is that it adds two columns including ex-Soviet countries in the estimation sample. The addition of controls does not change the patterns shown in Table 21.6, although omitting income as a control does lead to moderately significant negative effects in the GMM estimate on net Gini, and in both the OLS and GMM estimates for gross Gini. This suggests that there may be other forces correlated with GDP and democracy that influence inequality, such as some of the structural transformation variables we examine below. Social unrest is the only variable that has an effect on inequality that is significant in the gross Gini specifications, and our point estimates on democracy are roughly unchanged. The addition of ex-Soviet countries to our estimation sample results in smaller magnitudes of the effect of democracy on inequality, consistent with the idea that inequality went up in these countries following democratization.

We also found (but are not reporting to save space) that democracy does not have any significant effect on other measures of inequality. In particular, in Appendix A we show that, with updated data and our sample, democracy appears to have no effect on the log of industrial wages and explain why this result is different from those of Rodrik (1999).

We have also experimented with other estimates of the Gini using a panel with data every 5 years constructed from the World Income and Inequality Dataset. Controlling for indicators of type of concept used to calculate the Gini (i.e., disposable income, consumption and so on) as well as indicators for data quality, we found broadly similar results, though generally for smaller samples.

Overall, although some specifications do show a negative impact of democracy on inequality, particularly the gross Gini, there is no consistent and robust impact. This contrasts with our results on tax to GDP ratio (or the total government revenue to GDP ratio). Though this could be because of the lower quality of inequality data, it might also reflect some of the theoretical forces we have suggested in the previous section. We will turn to an investigation of some of these channels after looking at the relationship between democracy and structural transformation next.

21.5.3 Democracy and Structural Transformation

While our results above suggest that democracy has little net impact on inequality despite increasing taxation, some of the theoretical models we examined above suggest mechanisms by which democracy could affect inequality independently of government redistribution. (The lowering of barriers to entry, provision of public goods, and the expansion of market opportunities under democracy could be **offsetting** any redistribution accomplished by the fiscal system.) Therefore we examine the effect of democracy on economic structure and education.

Tables 21.8–21.10 look at the impact of democracy on various measures of structural transformation and public goods provision. We focus on the nonagricultural share of employment, nonagricultural share of value added, and secondary enrollment (which is a flow measure, thus better reflecting the effect of democracy on educational investments). Each table has two panels: the top one has the same structure as Table 21.2, whereas the bottom one is similar to Table 21.5 and shows the robustness of the results. Overall, we find significant effects of democracy on these measures of structural transformation.

For example, Tables 21.8 and 21.9 show some significant effects of democratization on the size of the nonagricultural sector.[30] Table 21.8 shows that democratization increases the (log of) nonagricultural share of employment, but this effect is generally only significant at the 10% level in the top panel, and is not completely robust to all exogenously imposed values of ρ in Columns 6–10. The bottom panel shows more consistent and significant estimates, but the coefficients differ substantially between the OLS and GMM estimators. Table 21.9, on the other hand shows that democratization increases the nonagricultural share of GDP. We find significant effects across OLS and most GMM specifications, imposing lower values for ρ, and with various sets of controls. The estimated magnitudes are plausible, with democracy increasing the nonagricultural employment by 4–11% and nonagricultural share of GDP by between 6% and 10% in the long run.

Table 21.10 shows a generally robust long-run effect of democratization on log secondary school enrollment. Although the coefficient magnitudes differ substantially between the GMM and OLS estimators, the long-run effect is uniformly positive and generally significant. Together with the taxation results, this suggests that one important economic change that democracies implement is to tax and provide public goods such as schooling. Our GMM specification in Column 3 of the top panel shows that democracy increases secondary enrollment by 67.6% in the long run, with an associated p-value of 0.07.[31]

[30] Bates and Block (2013) find that democratization significantly increased agricultural productivity in Africa, which may also be part of the process of structural change.

[31] The contrast of these results with Aghion et al. (2012), who find that democracy, as measured by the polity score, reduces primary school enrollment, is partly owing to their different sample, dependent variable, and econometric specification. Indeed, Aghion et al. (2012) estimate models without the lagged dependent variable and also include several additional variables on the right-hand side, most notably, military expenditure per capita (which is problematic since it is correlated with democracy, making it a potential "bad control"). They also focus on primary schooling, and according to our discussion above, democracy may have different effects on primary and secondary enrollment depending on the current level of education of the median voter.

Table 21.8 Effects of democratization on the log of the nonagricultural share of population

			GMM			Assuming AR(1) coefficient				
						$\rho = 0$	$\rho = 0.25$	$\rho = 0.5$	$\rho = 0.75$	$\rho = 1$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Democracy lagged	0.81	0.61⋆	1.86⋆	1.71⋆	1.66⋆	1.48	1.22	0.96	0.69⋆	0.43
	(1.74)	(0.33)	(0.95)	(0.92)	(0.86)	(1.63)	(1.18)	(0.75)	(0.39)	(0.38)
Dep. Var. lagged		0.83⋆⋆⋆	0.83⋆⋆⋆	0.83⋆⋆⋆	0.84⋆⋆⋆					
		(0.05)	(0.06)	(0.06)	(0.06)					
Observations	350	313	252	252	252	313	313	313	313	313
Countries	62	61	60	60	60	61	61	61	61	61
Number of moments			56	40	40					
Hansen p-value			0.33	0.12	0.07					
AR2 p-value			0.10	0.08	0.10					
Democracy changes	23	21	18	18	18	21	21	21	21	21
Long-run effect	0.81	3.59	10.79	9.91	10.09	1.48	1.62	1.91	2.77	.
p-Value	0.64	0.06	0.01	0.02	0.02	0.37	0.31	0.21	0.08	.

	Ex. GDP per capita		Adding other controls							
	OLS	GMM	OLS	GMM	OLS	GMM	OLS	GMM		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Democracy lagged	0.96⋆	1.34	0.70⋆	2.24⋆	0.68⋆	2.00⋆⋆	0.90⋆	1.60⋆⋆⋆	0.85⋆	1.38⋆⋆⋆
	(0.51)	(1.19)	(0.40)	(1.19)	(0.35)	(0.96)	(0.48)	(0.62)	(0.50)	(0.52)
Dep. Var. lagged	0.83⋆⋆⋆	0.84⋆⋆⋆	0.82⋆⋆⋆	0.81⋆⋆⋆	0.82⋆⋆⋆	0.81⋆⋆⋆	0.79⋆⋆⋆	0.79⋆⋆⋆	0.77⋆⋆⋆	0.77⋆⋆⋆
	(0.04)	(0.05)	(0.05)	(0.06)	(0.05)	(0.06)	(0.03)	(0.04)	(0.04)	(0.04)
War lagged			0.09	0.13					0.12	0.50⋆
			(0.08)	(0.17)					(0.12)	(0.28)
Unrest lagged					−0.00	0.01			−0.00	0.03⋆⋆
					(0.01)	(0.02)			(0.01)	(0.01)
Education lagged							−0.04	−0.06⋆	−0.03	−0.04
							(0.03)	(0.03)	(0.03)	(0.03)
Observations	341	279	229	184	294	237	227	183	189	153

Continued

Table 21.8 Effects of democratization on the log of the nonagricultural share of population—cont'd

	Ex. GDP per capita		Adding other controls							
	OLS	GMM	OLS	GMM	OLS	GMM	OLS	GMM		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Countries	62	61	45	44	57	56	44	43	36	35
Number of moments		55		57		57		52		54
Hansen p-value		0.29		0.81		0.31		0.92		1.00
AR2 p-value		0.82		0.57		0.21		0.23		0.29
Democracy changes	22	19	18	16	21	18	8	6	8	6
Long-run effect	5.72	8.42	3.79	11.87	3.84	10.63	4.20	7.52	3.77	5.90
p-Value	0.04	0.16	0.08	0.03	0.04	0.01	0.07	0.01	0.10	0.01

Notes for top panel: OLS estimates (Columns 1–2) include a full set of country and year fixed effects. Arellano and Bond's GMM estimators of the dynamic panel model (Columns 3–4) remove country fixed effects by taking first differences of the data, or by taking forward orthogonal differences (Column 5) and then constructing moment conditions using predetermined lags of the dependent variable and democracy. Columns 4 and 5 use up to the fifth lag of predetermined variables to create moments, restricting the number of moments used. Columns 6–10 impose different values for the autocorrelation coefficient in the percentage of nonagricultural population series, and estimate the effect of democracy including a full set of country and year fixed effects. Robust standard errors, adjusted for clustering at the country level, are in parentheses. *Notes for bottom panel*: OLS estimates (odd columns) include a full set of country and year fixed effects. Columns 3–10 include lagged GDP per capita as a control. Arellano and Bond's GMM estimators of the dynamic panel model (even columns) remove country fixed effects by taking first differences of the data and then constructing moment conditions using predetermined lags of the dependent variable and democracy. Robust standard errors, adjusted for clustering at the country level, are in parentheses. ***: significant at 1%; **: significant at 5%; *: significant at 10%. We do not report long-run effects and their p-values in Column 10 because they are not defined for $\rho = 1$.

Table 21.9 Effects of democratization on the log of nonagricultural share of GDP

			GMM			Assuming AR(1) coefficient				
						$\rho = 0$	$\rho = 0.25$	$\rho = 0.5$	$\rho = 0.75$	$\rho = 1$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Democracy lagged	3.96★★★	2.49★★★	2.66★	1.58	2.62★★	4.00★★★	3.34★★★	2.68★★★	2.02★★	1.36
	(1.38)	(0.95)	(1.56)	(2.15)	(1.31)	(1.34)	(1.11)	(0.95)	(0.92)	(1.01)
Dep. Var. lagged		0.57★★★	0.73★★★	0.74★★★	0.73★★★					
		(0.05)	(0.08)	(0.09)	(0.07)					
Observations	1033	978	833	833	834	978	978	978	978	978
Countries	147	144	140	140	140	144	144	144	144	144
Number of moments			100	70	70					
Hansen p-value			0.21	0.18	0.08					
AR2 p-value			0.72	0.71	0.40					
Democracy changes	90	88	78	78	78	88	88	88	88	88
Long-run effect	3.96	5.81	9.86	6.14	9.76	4.00	4.46	5.36	8.09	.
p-Value	0.00	0.01	0.10	0.49	0.08	0.00	0.00	0.01	0.03	.

	Ex. GDP per capita		Adding other controls							
	OLS	GMM	OLS	GMM	OLS	GMM	OLS	GMM		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Democracy lagged	2.63★★★	3.61★★	2.63★★	3.05★	2.43★★	2.67★	2.78★★★	3.70★★	2.82★★★	3.89★★
	(0.91)	(1.57)	(1.01)	(1.68)	(0.94)	(1.54)	(1.05)	(1.66)	(1.05)	(1.67)
Dep. Var. lagged	0.61★★★	0.78★★★	0.58★★★	0.73★★★	0.58★★★	0.75★★★	0.60★★★	0.76★★★	0.60★★★	0.74★★★
	(0.04)	(0.07)	(0.05)	(0.08)	(0.05)	(0.08)	(0.04)	(0.07)	(0.04)	(0.07)
War lagged			−0.45	−1.82★★					−0.29	−1.79★
			(0.44)	(0.74)					(0.43)	(0.96)
Unrest lagged					0.00	−0.01			0.00	−0.01
					(0.01)	(0.01)			(0.01)	(0.01)
Education lagged							−0.01	−0.15	0.01	−0.10
							(0.04)	(0.13)	(0.05)	(0.13)

Continued

Table 21.9 Effects of democratization on the log of nonagricultural share of GDP—cont'd

	Ex. GDP per capita		**Adding other controls**							
	OLS	**GMM**	**OLS**	**GMM**	**OLS**	**GMM**	**OLS**	**GMM**		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
Observations	1010	861	852	730	924	789	823	709	762	658
Countries	148	143	121	117	134	130	113	109	103	99
Number of moments		99		101		101		101		103
Hansen p-value		0.28		0.27		0.30		0.21		0.31
AR2 p-value		0.28		0.52		0.61		0.74		0.64
Democracy changes	91	81	81	74	88	78	70	63	69	62
Long-run effect	6.69	16.61	6.24	11.23	5.78	10.47	7.00	15.39	7.05	14.87
p-Value	0.01	0.08	0.01	0.09	0.01	0.10	0.01	0.02	0.01	0.02

Notes for top panel: OLS estimates (Columns 1–2) include a full set of country and year fixed effects. Arellano and Bond's GMM estimators of the dynamic panel model (Columns 3–4) remove country fixed effects by taking first differences of the data, or by taking forward orthogonal differences (Column 5) and then constructing moment conditions using predetermined lags of the dependent variable and democracy. Columns 4 and 5 use up to the fifth lag of predetermined variables to create moments, restricting the number of moments used. Columns 6–10 impose different values for the autocorrelation coefficient in the nonagricultural share of GDP series, and estimate the effect of democracy including a full set of country and year fixed effects. Robust standard errors, adjusted for clustering at the country level, are in parentheses. *Notes for bottom panel*: OLS estimates (odd columns) include a full set of country and year fixed effects. Columns 3–10 include lagged GDP per capita as a control. Arellano and Bond's GMM estimators of the dynamic panel model (even columns) remove country fixed effects by taking first differences of the data and then constructing moment conditions using predetermined lags of the dependent variable and democracy. Robust standard errors, adjusted for clustering at the country level, are in parentheses. ***: significant at 1%; **: significant at 5%; *: significant at 10%. We do not report long-run effects and their p-values in Column 10 because they are not defined for $\rho = 1$.

Table 21.10 Effects of democratization on the log of secondary enrollment

			GMM			Assuming AR(1) coefficient				
						$\rho = 0$	$\rho = 0.25$	$\rho = 0.5$	$\rho = 0.75$	$\rho = 1$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Democracy lagged	12.31★★	12.30★★★	17.41★★	20.35★★	13.39	19.28★★★	16.25★★★	13.22★★★	10.19★★	7.17
	(5.17)	(4.67)	(8.21)	(9.28)	(8.41)	(5.64)	(5.00)	(4.56)	(4.40)	(4.54)
Dep. Var. lagged		0.58★★★	0.74★★★	0.75★★★	0.82★★★					
		(0.06)	(0.12)	(0.12)	(0.12)					
Observations	825	630	453	453	489	630	630	630	630	630
Countries	150	141	127	127	129	141	141	141	141	141
Number of moments			77	56	57					
Hansen p-value			0.04	0.04	0.12					
AR2 p-value			0.83	0.91	0.79					
Democracy changes	71	51	29	29	29	51	51	51	51	51
Long-run effect	12.31	29.03	67.56	82.43	76.17	19.28	21.67	26.44	40.77	.
p-Value	0.02	0.01	0.07	0.09	0.16	0.00	0.00	0.00	0.02	.

	Ex. GDP per capita		Adding other controls							
	OLS	GMM	OLS	GMM	OLS	GMM	OLS	GMM		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Democracy lagged	11.19★★	16.77★★	12.85★★	21.47★★	12.81★★★	19.55★★	10.92★★	18.39★★	11.08★★	17.36★★
	(4.45)	(8.50)	(4.98)	(8.49)	(4.70)	(7.79)	(5.11)	(8.67)	(5.25)	(8.13)
Dep. Var. lagged	0.58★★★	0.80★★★	0.57★★★	0.72★★★	0.57★★★	0.73★★★	0.61★★★	0.82★★★	0.61★★★	0.81★★★
	(0.06)	(0.11)	(0.06)	(0.12)	(0.06)	(0.12)	(0.06)	(0.09)	(0.07)	(0.09)
War lagged			0.73	−0.07					0.31	−1.21
			(0.98)	(1.87)					(1.01)	(1.92)
Unrest lagged					0.04	0.04			0.06	0.12
					(0.04)	(0.06)			(0.05)	(0.07)
Education lagged							−0.29★	−0.78★	−0.32★	−0.74★
							(0.17)	(0.42)	(0.18)	(0.39)

Continued

Table 21.10 Effects of democratization on the log of secondary enrollment—cont'd

	Ex. GDP per capita		Adding other controls							
	OLS	**GMM**	**OLS**	**GMM**	**OLS**	**GMM**	**OLS**	**GMM**		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
Observations	686	495	563	411	610	442	553	407	519	385
Countries	151	134	121	111	133	121	116	106	106	99
Number of moments		76		78		78		78		80
Hansen p-value		0.08		0.13		0.04		0.19		0.18
AR2 p-value		0.66		0.67		0.61		0.51		0.59
Democracy changes	54	33	48	29	51	29	43	26	42	26
Long-run effect	26.50	84.72	30.11	76.17	30.04	71.71	28.20	103.65	28.32	93.58
p-Value	0.01	0.11	0.01	0.04	0.00	0.04	0.02	0.04	0.02	0.03

Notes for top panel: OLS estimates (Columns 1–2) include a full set of country and year fixed effects. Arellano and Bond's GMM estimators of the dynamic panel model (Columns 3–4) remove country fixed effects by taking first differences of the data, or by taking forward orthogonal differences (Column 5) and then constructing moment conditions using predetermined lags of the dependent variable and democracy. Columns 4 and 5 use up to the fifth lag of predetermined variables to create moments, restricting the number of moments used. Columns 6–10 impose different values for the autocorrelation coefficient in the secondary enrollment series, and estimate the effect of democracy including a full set of country and year fixed effects. Robust standard errors, adjusted for clustering at the country level, are in parentheses. *Notes for bottom panel*: OLS estimates (odd columns) include a full set of country and year fixed effects. Columns 3–10 include lagged GDP per capita as a control. Arellano and Bond's GMM estimators of the dynamic panel model (even columns) remove country fixed effects by taking first differences of the data and then constructing moment conditions using predetermined lags of of the dependent variable and democracy. Robust standard errors, adjusted for clustering at the country level, are in parentheses. ***: significant at 1%; **: significant at 5%; *: significant at 10%. We do not report long-run effects and their p-values in Column 10 because they are not defined for $\rho = 1$.

In sum, there is strong evidence that democratization does not just redistribute income, but also results in a degree of structural change of the economy and investment in public goods.[32] As our theoretical discussion implied, this could explain why democratization has a statistically weak effect on inequality. Democracy may be bringing new opportunities and economic change, which may increase inequality, while simultaneously lowering barriers to entry and investing in public goods, which may reduce inequality, and the net result could be either an increase or decrease in inequality, despite the increased taxation documented in Tables 21.2 and 21.3. This reasoning, as well as the theoretical ideas discussed in Section 21.2, underscores the importance of investigating the heterogeneous effects of democracy on inequality, a topic we turn to next.

21.5.4 Investigating the Mechanisms: Heterogeneity

We now turn to heterogeneity in the effect of democracy on inequality. We first consider the effect of democracy interacted with the land Gini, which we take to be a measure of landed elite power, to test the "capture" channel discussed above. We show only effects on net and gross Gini for most of the interactions to save space, and then discuss the heterogeneous effects on tax to GDP and government revenue to GDP ratios in the text.

Table 21.11 shows a positive and generally significant interaction of democracy with land inequality, suggesting that the power of landed elites to capture the state or thwart any redistributive tendencies of democratization results in higher inequality. The magnitudes are sizable, suggesting that a democratization in, say, Myanmar, with the highest land Gini (=77 in a 0- to 100-scale) among nondemocracies in our sample, would increase the after-tax Gini by approximately 0.72–2.42 points and the pretax Gini by 0.2–1.6 points. Our results suggest that democracy may increase inequality in societies with strong landed elites. This could be the case if democracy creates inequality increasing market opportunities while the elite manages to reduce taxation through de facto channels. An alternative explanation is given in Acemoglu and Robinson (2008), where a transition to democracy can lead to more pro-elite policies. The intuition for this somewhat paradoxical result is that the elite invests more in de facto power under democracy because, besides the benefits of being able to impose their favorite economic institutions, investments in de facto power increase the likelihood of a transition to autocracy.

The difference between the net and gross measures may reflect the importance of nonfiscal channels. Consistent with this, we see only moderate attenuation of the effect of democracy on the tax to GDP ratio, and no significant heterogeneity on the government revenue to GDP ratio (omitted to save space). For example, the equalizing effects of lowering barriers to mobility out of the agricultural sector may only be seen in societies

[32] Event study figures analogous to Figures 21.2, 21.4, and 21.5 reveal no pre-trends for these variables and an increase after the democratization, but are not included to save space.

Table 21.11 Effects of democratization on inequality (Includes interaction of democracy with Land Gini (averaged over all years with available data))

			GMM			Assuming AR(1) coefficient				
						$\rho = 0$	$\rho = 0.25$	$\rho = 0.5$	$\rho = 0.75$	$\rho = 1$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent variable: Gini coefficient, net income										
Democracy lagged	0.29	−0.91	−1.01	−2.44	−1.56	−0.56	−0.81	−1.06	−1.31	−1.57
	(0.90)	(1.02)	(1.31)	(1.86)	(1.46)	(1.04)	(1.02)	(1.05)	(1.12)	(1.22)
Lagged democracy × Land Gini	0.18***	0.11**	0.23**	0.33***	0.27***	0.15***	0.12**	0.09	0.06	0.03
	(0.04)	(0.05)	(0.10)	(0.10)	(0.10)	(0.05)	(0.05)	(0.06)	(0.07)	(0.08)
Dep. Var. lagged		0.35***	0.36***	0.33***	0.34***					
		(0.07)	(0.08)	(0.08)	(0.10)					
Observations	485	407	326	326	329	407	407	407	407	407
Countries	86	78	72	72	72	78	78	78	78	78
Democracy changes in the sample	32	23	16	16	16	23	23	23	23	23
Dependent variable: Gini coefficient, gross income										
Democracy lagged	−2.89**	−2.51**	−2.47*	−4.38**	−2.57**	−2.88*	−2.69**	−2.50**	−2.30**	−2.11**
	(1.32)	(1.04)	(1.28)	(2.07)	(1.29)	(1.48)	(1.21)	(1.03)	(0.96)	(1.05)
Lagged democracy × Land Gini	0.22***	0.18***	0.27***	0.46***	0.29***	0.24***	0.21***	0.18**	0.15**	0.12
	(0.08)	(0.07)	(0.08)	(0.12)	(0.07)	(0.09)	(0.08)	(0.07)	(0.07)	(0.07)
Dep. Var. lagged		0.48***	0.56***	0.49***	0.66***					
		(0.07)	(0.09)	(0.11)	(0.09)					
Observations	485	407	326	326	329	407	407	407	407	407
Countries	86	78	72	72	72	78	78	78	78	78
Democracy changes in the sample	32	23	16	16	16	23	23	23	23	23

Note: OLS estimates (Columns 1–2) include a full set of country and year fixed effects. Arellano and Bond's GMM estimators of the dynamic panel model (Columns 3–4) remove country fixed effects by taking first differences of the data, or by taking forward orthogonal differences (Column 5) and then constructing moment conditions using predetermined lags of the dependent variable, the interaction term and democracy. Columns 4 and 5 use up to the fifth lag of predetermined variables to create moments, restricting the number of moments used. Columns 6–10 impose different values for the autocorrelation coefficient in the tax to GDP series, and estimate the effect of democracy and the interaction term including a full set of country and year fixed effects. All models control for lagged GDP per capita but this coefficient is not reported to save space. Robust standard errors, adjusted for clustering at the country level, are in parentheses. ***: significant at 1%; **: significant at 5%; *: significant at 10%.

with politically weak agricultural elites. Although land inequality is potentially correlated with many other economic and social factors that may also mediate the effect of democracy on inequality, we view this as some evidence of the "capture" channel modeled above.

We next consider the effect of democracy depending on the extent of structural transformation, motivated by our hypothesis that democracy induces structural change and may increase inequality by expanding opportunities, such as skilled occupations and entrepreneurship, for previously excluded groups.

Table 21.12 shows the effect of democratization interacted with the share of nonagricultural employment in 1968 as a measure of the extent of structural transformation (results are similar with the 1978 share). We find that democratization *increases* inequality more (or fails to reduce inequality) in places that have smaller agricultural employment shares. This is consistent with democracy expanding access to inequality-increasing market opportunities especially in more urban societies where skilled occupations and entrepreneurship are potentially more important. The magnitudes suggest that democratization in a country that was 10% points less agricultural than the mean in 1968 (measured by the percentage of nonagricultural employment), will bring an increase between 1 and 1.6 net Gini points (1.3 and 2.3 gross Gini points) relative to the average effect in the short run, and between 1.6 and 2.2 net Gini points (2.5 and 5.6 gross Gini points) in the long run. We have also estimated these specifications using our other proxies for structural transformation and obtained uniformly positive, although often imprecise, coefficients on the interaction variables. The results using the gross Gini coefficient show a similar pattern and similar, though slightly larger, estimates.

While we do not show these results for space reasons, there is no significant heterogeneity by nonagricultural employment in the effect of democracy on taxation, and this result is robust to all proxies for the extent of structural transformation we have tried, including the 1970 values of urbanization, education, and nonagricultural share of GDP. This suggests that the mechanisms via which democracy increases inequality in relatively more economically modernized countries has less to do with lowering government redistribution or public good provision, and more to do with other mechanisms emphasized in our discussion of disequalizing market opportunities opened up by democracy for entrepreneurs, educated workers, and capitalists.

Table 21.13 looks further at heterogeneity by the level of potential inequality created by market opportunities. We interact democratization in year t with the top 10% share of income in the United States in the same year. This is a proxy (albeit a highly imprecise and imperfect one) for the extent of inequality increasing market opportunities available at the time and their potential to create inequality, shaped by world-level forces such as globalization, technological and organizational changes that either originate or find widespread adoption in the United States (Panitch and Gindin, 2012). We did not find significant interaction effects of this sort on the tax to GDP ratio or the government revenue

Table 21.12 Effects of democratization on inequality (Includes interaction of democracy with the percentage of nonagricultural population in 1968)

			GMM			Assuming AR(1) coefficient				
						$\rho = 0$	$\rho = 0.25$	$\rho = 0.5$	$\rho = 0.75$	$\rho = 1$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent variable: Gini coefficient, net income										
Democracy lagged	0.91	−0.32	−0.45	−1.81	−0.80	−0.05	−0.27	−0.49	−0.71	−0.92
	(0.74)	(0.78)	(1.35)	(1.54)	(1.28)	(0.82)	(0.79)	(0.81)	(0.88)	(0.98)
Lagged democracy × nonagricultural	0.12⋆⋆⋆	0.11⋆⋆	0.16⋆	0.16⋆⋆	0.13⋆	0.13⋆⋆⋆	0.12⋆⋆	0.10⋆⋆	0.09	0.08
pop. in 1968	(0.03)	(0.05)	(0.08)	(0.07)	(0.07)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)
Dep. Var. lagged		0.31⋆⋆⋆	0.29⋆⋆⋆	0.29⋆⋆⋆	0.36⋆⋆⋆					
		(0.07)	(0.09)	(0.10)	(0.10)					
Observations	614	506	402	402	406	506	506	506	506	506
Countries	112	100	91	91	91	100	100	100	100	100
Democracy changes in the sample	55	41	29	29	29	41	41	41	41	41
Dependent variable: Gini coefficient, gross income										
Democracy lagged	−0.81	−0.85	−0.40	−1.15	−0.71	−0.72	−0.79	−0.86	−0.92	−0.99
	(0.98)	(0.76)	(1.22)	(1.43)	(1.18)	(0.97)	(0.83)	(0.76)	(0.79)	(0.90)
Lagged democracy × nonagricultural	0.15⋆⋆⋆	0.13⋆⋆	0.21⋆⋆⋆	0.23⋆⋆⋆	0.19⋆⋆⋆	0.17⋆⋆⋆	0.15⋆⋆⋆	0.13⋆⋆	0.11⋆⋆	0.08⋆
pop. in 1968	(0.05)	(0.05)	(0.08)	(0.08)	(0.07)	(0.07)	(0.06)	(0.05)	(0.05)	(0.04)
Dep. Var. lagged		0.48⋆⋆⋆	0.54⋆⋆⋆	0.55⋆⋆⋆	0.66⋆⋆⋆					
		(0.06)	(0.10)	(0.10)	(0.10)					
Observations	614	506	402	402	406	506	506	506	506	506
Countries	112	100	91	91	91	100	100	100	100	100
Democracy changes in the sample	55	41	29	29	29	41	41	41	41	41

Note: OLS estimates (Columns 1–2) include a full set of country and year fixed effects. Arellano and Bond's GMM estimators of the dynamic panel model (Columns 3–4) remove country fixed effects by taking first differences of the data, or by taking forward orthogonal differences (Column 5) and then constructing moment conditions using predetermined lags of the dependent variable, the interaction term and democracy. Columns 4 and 5 use up to the fifth lag of predetermined variables to create moments, restricting the number of moments used. Columns 6–10 impose different values for the autocorrelation coefficient in the tax to GDP series, and estimate the effect of democracy and the interaction term including a full set of country and year fixed effects. All models control for lagged GDP per capita but this coefficient is not reported to save space. Robust standard errors, adjusted for clustering at the country level, are in parentheses. ⋆⋆⋆: significant at 1%; ⋆⋆: significant at 5%; ⋆: significant at 10%.

Table 21.13 Effects of democratization on inequality (Includes interaction of democracy with share of income held by the top 10 decile in the United States at the time of democratization)

			GMM			Assuming AR(1) coefficient				
						$\rho=0$	$\rho=0.25$	$\rho=0.5$	$\rho=0.75$	$\rho=1$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent variable: Gini coefficient, net income										
Democracy lagged	0.68	−0.76	−2.35	−3.06⋆	−0.88	−0.46	−0.70	−0.94	−1.18	−1.42
	(0.79)	(0.89)	(1.57)	(1.64)	(1.52)	(0.94)	(0.90)	(0.90)	(0.93)	(1.00)
Lagged democracy × Top 10 share in the US	0.22⋆	0.19⋆	−0.10	−0.12	0.22	0.27⋆	0.21⋆	0.14	0.08	0.01
	(0.13)	(0.11)	(0.19)	(0.19)	(0.17)	(0.15)	(0.12)	(0.10)	(0.09)	(0.10)
Dep. Var. lagged		0.31⋆⋆⋆	0.35⋆⋆⋆	0.36⋆⋆⋆	0.47⋆⋆⋆					
		(0.07)	(0.10)	(0.11)	(0.08)					
Observations	657	537	420	420	424	537	537	537	537	537
Countries	127	113	100	100	100	113	113	113	113	113
Democracy changes in the sample	65	47	31	31	31	47	47	47	47	47
Dependent variable: Gini coefficient, gross income										
Democracy lagged	−1.04	−1.55	−0.68	−0.71	0.44	−1.61	−1.58	−1.54	−1.51⋆	−1.48
	(0.98)	(0.95)	(1.46)	(1.76)	(1.69)	(1.18)	(1.04)	(0.94)	(0.89)	(0.91)
Lagged democracy × Top 10 share in the US	0.72⋆⋆⋆	0.36⋆⋆⋆	0.25	0.28	0.37⋆⋆	0.72⋆⋆⋆	0.52⋆⋆⋆	0.33⋆⋆⋆	0.13	−0.06
	(0.17)	(0.11)	(0.18)	(0.21)	(0.16)	(0.19)	(0.15)	(0.11)	(0.09)	(0.10)
Dep. Var. lagged		0.46⋆⋆⋆	0.61⋆⋆⋆	0.60⋆⋆⋆	0.71⋆⋆⋆					
		(0.06)	(0.11)	(0.11)	(0.11)					
Observations	657	537	420	420	424	537	537	537	537	537
Countries	127	113	100	100	100	113	113	113	113	113
Democracy changes in the sample	65	47	31	31	31	47	47	47	47	47

Note: OLS estimates (Columns 1–2) include a full set of country and year fixed effects. Arellano and Bond's GMM estimators of the dynamic panel model (Columns 3–4) remove country fixed effects by taking first differences of the data, or by taking forward orthogonal differences (Column 5) and then constructing moment conditions using predetermined lags of the dependent variable, the interaction term and democracy. Columns 4 and 5 use up to the fifth lag of predetermined variables to create moments, restricting the number of moments used. Columns 6–10 impose different values for the autocorrelation coefficient in the tax to GDP series, and estimate the effect of democracy and the interaction term including a full set of country and year fixed effects. All models control for lagged GDP per capita but this coefficient is not reported to save space. Robust standard errors, adjusted for clustering at the country level, are in parentheses. ⋆⋆⋆: significant at 1%; ⋆⋆: significant at 5%; ⋆: significant at 10%.

to GDP ratio. However, we do see generally significant impact of this interaction on the gross Gini, which appears to be further increased by democracy when there is greater inequality in the United States. There is also a similar effect on the net Gini but is much weaker and not present when using the GMM estimators. Though on the whole this evidence is on the weak side, it is broadly consistent with a story in which democratization increases inequality at times when the expanded market opportunities available are more disequalizing.

Finally, Tables 21.14–21.17 provide some preliminary evidence on Director's law. Recall from our discussion in Section 21.2, in particular Proposition 5, that our (modified) Director's law implies that the negative effect of democracy on inequality should be visible or greater in places where the rich have a large share of income (Meltzer-Richards also predicts this) and, more uniquely, should be positive where the poor have a higher share of income (which is the opposite of the Meltzer-Richards prediction). Thus, we investigate the heterogeneous effect of democracy depending on the shares of the top and bottom of the income distribution (in each case relative to the share of the middle, i.e., using the top and bottom income shares described above). Recall also that the effect of the income share of the rich on inequality in democracy is related to whether there is capture of democracy by the elite, which provides a reason why this prediction of Proposition 5 may not hold even when a greater share of income of the poor may increase inequality as posited in Proposition 5.

Indeed, Table 21.14 shows that when the top decile is richer relative to the middle, there is no significantly heterogeneous effect on inequality, although coefficients are generally negative. This might be because this estimate is picking up both an elite capture effect (as in the land Gini interaction specifications) as well as additional demand for redistribution by the median voter as in our (modified) Director's law, with higher incidence on the rich. Table 21.15 provides support for the possibility that top tail inequality, as measured by the top share, could be picking up elite capture effects. It shows that the effect of democracy on the tax to GDP ratio is significantly attenuated by income inequality as measured by the top share (but there is no effect on government revenue as a fraction of GDP), contrary to what Meltzer and Richards model or our (modified) Director's law would predict. Our conclusion from this exercise is that our research design does not allow us to separate the effects of democracy through the demand for redistribution and the incidence of taxation emphasized in our modified Director's law from the possibility that democracies with large upper tail inequality are more likely to be captured by the wealthier elite.

Tables 21.16 and 21.17, on the other hand, provide support for the more unique prediction of the (modified) Director's law, that democracy should increase inequality more when the poor are closer to the middle class in nondemocracy. Table 21.16 looks at the interaction of the bottom income share with democracy, and finds that the net Gini does in fact increase with democratization, while there is no effect on the gross Gini. This

Table 21.14 Effects of democratization on inequality (Includes interaction of democracy with the average share of income held by the top decile relative to share of mid 50th earners before 2000)

			GMM			Assuming AR(1) coefficient				
						$\rho=0$	$\rho=0.25$	$\rho=0.5$	$\rho=0.75$	$\rho=1$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent variable: Gini coefficient, net income										
Democracy lagged	0.79	−0.54	−1.39	−1.73	−1.24	−0.20	−0.48	−0.76	−1.04	−1.32
	(0.80)	(0.88)	(1.47)	(1.49)	(1.33)	(0.95)	(0.91)	(0.91)	(0.95)	(1.03)
Lagged democracy × Top share	−0.01	−0.01	−0.02	−0.02	−0.01	−0.01	−0.01	−0.01	−0.01	−0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Dep. Var. lagged		0.30⋆⋆⋆	0.29⋆⋆⋆	0.30⋆⋆⋆	0.32⋆⋆⋆					
		(0.07)	(0.08)	(0.08)	(0.08)					
Observations	606	503	397	397	401	503	503	503	503	503
Countries	110	102	93	93	93	102	102	102	102	102
Democracy changes in the sample	55	41	29	29	29	41	41	41	41	41
Dependent variable: Gini coefficient, gross income										
Democracy lagged	−0.76	−1.29	−1.73	−2.30	−1.55	−1.02	−1.16	−1.30	−1.45	−1.59
	(0.93)	(0.85)	(1.31)	(1.41)	(1.30)	(0.98)	(0.89)	(0.85)	(0.88)	(0.97)
Lagged democracy × Top share	−0.00	−0.00	0.00	−0.00	0.00	−0.00	−0.00	−0.00	−0.01	−0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Dep. Var. lagged		0.48⋆⋆⋆	0.52⋆⋆⋆	0.54⋆⋆⋆	0.60⋆⋆⋆					
		(0.06)	(0.08)	(0.08)	(0.08)					
Observations	606	503	397	397	401	503	503	503	503	503
Countries	110	102	93	93	93	102	102	102	102	102
Democracy changes in the sample	55	41	29	29	29	41	41	41	41	41

Note: OLS estimates (Columns 1–2) include a full set of country and year fixed effects. Arellano and Bond's GMM estimators of the dynamic panel model (Columns 3–4) remove country fixed effects by taking first differences of the data, or by taking forward orthogonal differences (Column 5) and then constructing moment conditions using predetermined lags of the dependent variable, the interaction term and democracy. Columns 4 and 5 use up to the fifth lag of predetermined variables to create moments, restricting the number of moments used. Columns 6–10 impose different values for the autocorrelation coefficient in the tax to GDP series, and estimate the effect of democracy and the interaction term including a full set of country and year fixed effects. All models control for lagged GDP per capita but this coefficient is not reported to save space. Robust standard errors, adjusted for clustering at the country level, are in parentheses. ⋆⋆⋆: significant at 1%; ⋆⋆: significant at 5%; ⋆: significant at 10%.

Table 21.15 Effects of democratization on the log of tax and total government revenue as a percentage of GDP (Includes interaction of democracy with the average share of income held by the top decile relative to share of mid 50th earners before 2000)

			GMM			Assuming AR(1) coefficient				
						$\rho=0$	$\rho=0.25$	$\rho=0.5$	$\rho=0.75$	$\rho=1$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent variable: Tax revenues as a percentage of GDP										
Democracy lagged	18.75***	14.54***	20.93***	21.97**	19.86**	18.75***	14.50***	10.24***	5.99**	1.74
	(4.88)	(3.72)	(8.02)	(9.86)	(8.55)	(4.88)	(3.46)	(2.48)	(2.52)	(3.54)
Lagged democracy × Top share	−0.10***	−0.08***	−0.22**	−0.19**	−0.20**	−0.10***	−0.08***	−0.06***	−0.03*	−0.01
	(0.04)	(0.03)	(0.09)	(0.08)	(0.10)	(0.04)	(0.03)	(0.02)	(0.02)	(0.02)
Dep. Var. lagged		0.25***	0.23***	0.25***	0.30***					
		(0.06)	(0.09)	(0.08)	(0.08)					
Observations	843	843	730	730	730	843	843	843	843	843
Countries	113	113	110	110	110	113	113	113	113	113
Democracy changes in the sample	72	72	67	67	67	72	72	72	72	72
Dependent variable: Total government revenues as a percentage of GDP										
Democracy lagged	10.56**	8.46***	14.27**	15.50**	13.97**	10.56**	9.09***	7.61***	6.13***	4.66
	(4.03)	(2.43)	(6.11)	(7.07)	(6.86)	(4.03)	(2.81)	(2.02)	(2.19)	(3.17)
Lagged democracy × Top share	−0.03	−0.02	−0.10	−0.11	−0.12	−0.03	−0.02	−0.02	−0.01	−0.00
	(0.03)	(0.02)	(0.06)	(0.07)	(0.08)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Dep. Var. lagged		0.36***	0.43***	0.48***	0.50***					
		(0.04)	(0.06)	(0.06)	(0.06)					
Observations	843	843	730	730	730	843	843	843	843	843
Countries	113	113	110	110	110	113	113	113	113	113
Democracy changes in the sample	72	72	67	67	67	72	72	72	72	72

Note: OLS estimates (Columns 1–2) include a full set of country and year fixed effects. Arellano and Bond's GMM estimators of the dynamic panel model (Columns 3–4) remove country fixed effects by taking first differences of the data, or by taking forward orthogonal differences (Column 5) and then constructing moment conditions using predetermined lags of the dependent variable, the interaction term and democracy. Columns 4 and 5 use up to the fifth lag of predetermined variables to create moments, restricting the number of moments used. Columns 6–10 impose different values for the autocorrelation coefficient in the tax to GDP series, and estimate the effect of democracy and the interaction term including a full set of country and year fixed effects. All models control for lagged GDP per capita but this coefficient is not reported to save space. Robust standard errors, adjusted for clustering at the country level, are in parentheses. ***: significant at 1%; **: significant at 5%; *: significant at 10%.

Table 21.16 Effects of democratization on inequality (Includes interaction of democracy with the average share of income held by the bottom decile relative to share of mid 50th earners before 2000)

			GMM			Assuming AR(1) coefficient				
						$\rho=0$	$\rho=0.25$	$\rho=0.5$	$\rho=0.75$	$\rho=1$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent variable: Gini coefficient, net income										
Democracy lagged	0.92	−0.41	−2.11⋆	−2.64⋆	−1.93⋆	−0.07	−0.35	−0.64	−0.92	−1.21
	(0.78)	(0.85)	(1.28)	(1.35)	(1.14)	(0.91)	(0.88)	(0.88)	(0.93)	(1.01)
Lagged democracy × Bottom share	0.52⋆	0.67⋆⋆	0.94⋆	0.71	0.58	0.71⋆⋆	0.68⋆⋆	0.65⋆⋆	0.62⋆⋆	0.59⋆⋆
	(0.29)	(0.28)	(0.56)	(0.51)	(0.44)	(0.30)	(0.28)	(0.27)	(0.28)	(0.30)
Dep. Var. lagged		0.30⋆⋆⋆	0.33⋆⋆⋆	0.35⋆⋆⋆	0.43⋆⋆⋆					
		(0.07)	(0.07)	(0.08)	(0.07)					
Observations	606	503	397	397	401	503	503	503	503	503
Countries	110	102	93	93	93	102	102	102	102	102
Democracy changes in the sample	55	41	29	29	29	41	41	41	41	41
Dependent variable: Gini coefficient, gross income										
Democracy lagged	−0.68	−1.24	−1.57	−2.35⋆	−1.58	−0.95	−1.10	−1.25	−1.40	−1.55
	(0.95)	(0.86)	(1.26)	(1.42)	(1.24)	(1.00)	(0.91)	(0.87)	(0.89)	(0.97)
Lagged democracy × Bottom share	0.29	0.28	−0.08	−0.24	0.09	0.31	0.29	0.28	0.26	0.25
	(0.37)	(0.31)	(0.52)	(0.56)	(0.49)	(0.37)	(0.33)	(0.31)	(0.32)	(0.35)
Dep. Var. lagged		0.48⋆⋆⋆	0.58⋆⋆⋆	0.60⋆⋆⋆	0.66⋆⋆⋆					
		(0.06)	(0.08)	(0.08)	(0.07)					
Observations	606	503	397	397	401	503	503	503	503	503
Countries	110	102	93	93	93	102	102	102	102	102
Democracy changes in the sample	55	41	29	29	29	41	41	41	41	41

Note: OLS estimates (Columns 1–2) include a full set of country and year fixed effects. Arellano and Bond's GMM estimators of the dynamic panel model (Columns 3–4) remove country fixed effects by taking first differences of the data, or by taking forward orthogonal differences (Column 5) and then constructing moment conditions using predetermined lags of the dependent variable, the interaction term and democracy. Columns 4 and 5 use up to the fifth lag of predetermined variables to create moments, restricting the number of moments used. Columns 6–10 impose different values for the autocorrelation coefficient in the tax to GDP series, and estimate the effect of democracy and the interaction term including a full set of country and year fixed effects. All models control for lagged GDP per capita but this coefficient is not reported to save space. Robust standard errors, adjusted for clustering at the country level, are in parentheses. ⋆⋆⋆: significant at 1%; ⋆⋆: significant at 5%; ⋆: significant at 10%.

Table 21.17 Effects of democratization on the log of tax and total government revenue as a percentage of GDP (Includes interaction of democracy with the average share of income held by the bottom decile relative to share of mid 50th earners before 2000)

			GMM			Assuming AR(1) coefficient				
						$\rho=0$	$\rho=0.25$	$\rho=0.5$	$\rho=0.75$	$\rho=1$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent variable: Tax revenues as a percentage of GDP										
Democracy lagged	18.72***	14.44***	18.47*	22.49**	17.16**	18.72***	14.42***	10.11***	5.81**	1.50
	(5.18)	(3.98)	(9.43)	(10.35)	(8.34)	(5.18)	(3.70)	(2.61)	(2.49)	(3.45)
Lagged democracy × Bottom share	5.04***	3.88***	7.34	9.31*	6.36	5.04***	3.87***	2.70**	1.54	0.37
	(1.88)	(1.46)	(4.84)	(5.58)	(5.54)	(1.88)	(1.44)	(1.14)	(1.12)	(1.38)
Dep. Var. lagged		0.25***	0.24***	0.24***	0.30***					
		(0.06)	(0.09)	(0.08)	(0.07)					
Observations	843	843	730	730	730	843	843	843	843	843
Countries	113	113	110	110	110	113	113	113	113	113
Democracy changes in the sample	72	72	67	67	67	72	72	72	72	72
Dependent variable: Total government revenues as a percentage of GDP										
Democracy lagged	10.78***	8.30***	13.23*	14.20**	11.74*	10.78***	9.03***	7.29***	5.54**	3.80
	(4.07)	(2.46)	(6.91)	(7.05)	(7.10)	(4.07)	(2.84)	(2.01)	(2.12)	(3.08)
Lagged democracy × Bottom share	1.55	0.74	2.92	4.14	3.37	1.55	0.98	0.40	−0.17	−0.74
	(1.63)	(1.21)	(3.09)	(3.86)	(3.77)	(1.63)	(1.29)	(1.10)	(1.14)	(1.39)
Dep. Var. lagged		0.35***	0.43***	0.47***	0.50***					
		(0.04)	(0.06)	(0.06)	(0.05)					
Observations	843	843	730	730	730	843	843	843	843	843
Countries	113	113	110	110	110	113	113	113	113	113
Democracy changes in the sample	72	72	67	67	67	72	72	72	72	72

Note: OLS estimates (Columns 1–2) include a full set of country and year fixed effects. Arellano and Bond's GMM estimators of the dynamic panel model (Columns 3–4) remove country fixed effects by taking first differences of the data, or by taking forward orthogonal differences (Column 5) and then constructing moment conditions using predetermined lags of the dependent variable, the interaction term and democracy. Columns 4 and 5 use up to the fifth lag of predetermined variables to create moments, restricting the number of moments used. Columns 6–10 impose different values for the autocorrelation coefficient in the tax to GDP series, and estimate the effect of democracy and the interaction term including a full set of country and year fixed effects. All models control for lagged GDP per capita but this coefficient is not reported to save space. Robust standard errors, adjusted for clustering at the country level, are in parentheses. ***: significant at 1%; **: significant at 5%; *: significant at 10%.

relative difference between the pre-fiscal and post-fiscal effects suggests that government redistribution may be (part of) the mechanism. Table 21.17 confirms this by showing that the tax to GDP ratio does go up following a democratization in a society where the poor are initially relatively well-off compared to the middle class.

Subject to the major caveats about omitted variables and measurement error, this evidence thus provides some support to our (modified) Director's law: middle classes empowered by democracy appear to be able to use the government to transfer resources from the poor to themselves, increasing post-fiscal inequality. As far as we know, this is the first evidence of this kind on how democracy might redistribute in a way that *increases* inequality.

We have investigated a number of other sources of heterogeneity, including various measures of ethnolinguistic fragmentation, wheat-sugar land suitability ratio (as a measure of the type of agriculture), constitutional provisions against redistribution, and average level of social unrest, and found no robust results.

Overall, the important concerns about endogeneity and measurement error notwithstanding, the results presented in this section paint a picture in which democracy does indeed create greater pressures for redistribution, but the pathways via which these affect inequality are more nuanced than the standard Meltzer-Richard mechanism presumes. In particular, the correlation between democracy and inequality appears to be more limited than one might have at first expected (and more limited than the effect on taxes). On the other hand, the evidence on heterogeneity of effects, even if not as robustly estimated as the impact on taxes, indicates that interactions with elite capture, structural transformation, middle-class bias in redistribution, and the disequalizing market opportunities opened up by democracy might be playing some role in modulating the influence of democracy on inequality.

21.6. CONCLUSION

The effect of democracy on redistribution and inequality is important for understanding how democracies function and use the available policy instruments. Nevertheless, our survey of the relevant literature shows that the social science literature on this topic is far from a consensus or a near-consensus on this topic.

We explained why the baseline expectation in the literature has been that democracy should increase redistribution and reduce inequality (for example, based on Meltzer and Richard's, 1981 seminal paper), and why this expectation may not be borne out in the data because democracy may be captured or constrained; because democracy may cater to the wishes of the middle class; or because democracy may simultaneously open up new economic opportunities to the previously excluded, contributing to economic inequality. This ambiguity may be one of the reasons why the large empirical literature on this topic comes to such inconclusive findings, though the use of datasets with different qualities and different methodologies and econometric practices, many of which are far from satisfactory, are also contributing factors. It may also be that because different researchers have looked at different sets of countries in different periods, the differing results are to some extent picking up situations where one or another of the mechanisms we have identified is more dominant.

The bulk of the chapter empirically investigated the (dynamic) relationship between democracy and various economic outcomes related to redistribution and inequality. Our results, which come from panel data models controlling for the dynamics and persistence in our outcome variables, indicate that democratization does indeed increase government taxation and revenue as fractions of GDP. This confirms the basic prediction of the standard Meltzer-Richard model. In contrast, we have found no robust evidence that democracy reduces inequality, although our estimated coefficients are quite imprecise in this case. Our results also suggest that democracy increases the share of GDP and population not in agriculture, as well as secondary school enrollment. This is consistent with democracy triggering a more rapid structural transformation, for example, because this structural transformation may have been arrested or slowed down by the nondemocratic political system. The relationship between democratic institutions and structural change is worth further investigation.

These patterns suggest that the effect of democracy on redistribution and inequality may be more nuanced than often presumed and highly heterogeneous across societies. We tried to make some tentative progress on this issue by providing additional correlations pertaining to these heterogeneous effects and mechanisms on which they might be based. We found some results suggesting that democratization in the presence of powerful landed elites may increase inequality, and that structural transformation may induce an expansion of opportunities that counteract any additional redistribution, and either of these could explain the absence of an effect on inequality. This interpretation is confirmed by our finding that democracy increases inequality in places that have a lower share of population in agriculture, and at times when the global technological and organizational frontier is more inequality inducing. A natural next step for research is isolating exogenous variation in these heterogeneous effects across democracies and nondemocracies.

In addition, we also found some evidence consistent with a (modified) Director's law, which suggests that democracy redistributes from the rich and the poor to the middle class, and therefore its effect on inequality may depend on the relative position of the middle class vis-a-vis the poor and the rich. Further research on whether and how democracies transfer from the poor to the middle class would be an important contribution.

(Overall, the evidence suggests that to the impact of democracy on inequality is limited, and these limited effects work by altering pre-redistribution market outcomes, while the fiscal mechanisms stressed by the literature play at most a small role in explaining any effect of democracy on inequality, and may in fact be inequality-increasing. We hope that further research on these issues, tackling the first-order endogeneity concerns and exploiting within-country as well as cross-national variation, will more systematically uncover the mechanisms at work.)

ACKNOWLEDGMENTS

We are grateful to the editors for their detailed comments on an earlier draft and to participants in the Handbook conference in Paris, particularly to our discussant José-Víctor Ríos-Rull.

APPENDIX A. COMPARISON TO RODRIK (1999)

This appendix replicates and extends the analysis in Rodrik (1999). At a first glance, the fact that we find no robust effect on net or gross income inequality seems at odds with Rodrik's findings that democracies pay higher real wages in manufacturing. These opposite findings could be explained by a logic similar to the one outlined in Proposition 4. In particular, democracies may increase wages by allowing workers to reallocate to new sectors, but this may also increase inequality if there is sufficient heterogeneity in labor productivity and wages were previously compressed and reduced by labor market institutions. Besides this conceptual difference we also explore the differences between our empirical setting and Rodrik's. We show that while the results are robust to our democracy measure, they are fragile in a number of other directions.

Rodrik's data generating model is given by

$$\log w_{it} = \beta D_{it} + X_{it}\gamma + \delta_i + \delta_t + \Sigma_{it},$$

with w_{it} manufacturing wages from the UNIDO dataset compiled by Martin Rama. However, this model cannot be estimated because wage data comes grouped on averages for the years $t, t+1, t+2, t+3, t+4$ for every 5 years from 1960 onward. Thus, only the average wages between 1960 and 1964, 1965 and 1969, and so on are observed. Thus, Rodrik estimates

$$\log w_{it,t+4} = \beta D_{it,t+4} + X_{it,t+4}\gamma + \delta_i + \delta_t + \epsilon_{it,t+4}. \tag{21.A1}$$

with all variables averaged over 5 year periods (from t to $t+4$), and the model is estimated in a panel covering $1960, 1965, \ldots, 1990$. Though Rodrik presents cross-sectional and panel estimates, we focus on the latter which are the more convincing ones and are also closer to the empirical strategy adopted in this chapter.

In the top panel of Table 21.A1 we present different estimates of Equation (21.A1) using a normalized polity score between 0 and 1, a normalized Freedom House index between 0 and 1 and our democracy measure separately as proxies for democracy. We always control for the log of GDP per capita, the log of worker value added in

Table 21.A1 Replication of Rodrik's results on the log of manufacturing wages

	Original wage data			Updated wage data		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Averaging democracy measure over *t*, *t*+4						
Polity index at *t*, *t*+4	19.25*** (5.72)			14.48** (6.00)		
Freedom house index at *t*, *t*+4		15.78** (7.55)			7.60 (8.68)	
Our democracy index at *t*, *t*+4			8.48** (3.66)			6.51 (4.20)
Observations	442	365	468	451	364	467
Countries	93	98	99	90	92	92
Using democracy measure at *t*						
Polity index at *t*	8.40 (6.15)			9.01 (5.89)		
Freedom house index at *t*		11.03 (10.55)			11.52 (9.77)	
Our democracy index at *t*			1.98 (3.54)			2.89 (3.39)
Observations	429	285	455	437	294	456
Countries	91	96	97	85	87	90

Dependent variable is log of average wages between *t* and *t*+4.
Note: OLS estimates include a full set of country and year fixed effects. All models control for the log of GDP per capita, log of worker value added and log of the price level, but these coefficients are not reported to save space. Robust standard errors, adjusted for clustering at the country level, are in parentheses. ***: significant at 1%; **: significant at 5%; *: significant at 10%.

manufacturing and the log of the price index (from the Penn World Tables) following Rodrik's original setup. The estimates of β are multiplied by 100 to ease their interpretation. The left panel uses Rodrik's original wage data and the right panel uses an updated version. In all models we present robust standard errors adjusting for clustering at the country level, which are reflected in slightly higher standard errors than the ones found by Rodrik.

Our estimates show that democracy, measured by any of the indices, is associated with higher wages using the original wage data, which replicates Rodrik's findings. There are still some small differences caused by updates to Polity and Freedom House, but qualitatively his conclusions hold. In particular, an increase in the polity score from 0 to 1 increases wages by 19.72% (s.e. = 5.98); an increase in the Freedom House index from 0 to 1 increases wages by 20.57% (s.e. = 8.13), and a switch from nondemocracy to democracy in our measure increases wages by 8.54% (s.e. = 3.88). The results using the new wage data are less clear, smaller, and not significant for Freedom House and our democracy measure. The results suggest that the association between democracy and wages is not robust if one uses the updated wage data and the same empirical strategy as Rodrik.

There are two more issues that are important to consider in weighing the importance of Rodrik's evidence. The wage data are in the form of 5-year averages. First, this will tend to induce nontrivial serial correlation in the dependent variable, inducing error in the presence of lagged dependent variables on the right-hand side (which our estimates suggest are present). Second, by averaging the democracy index, Rodrik's specification induces the correlation between wages at t and democracy at $t+1, t+2, t+3$ and $t+4$, which of course does not reflect the effect of democracy on wages, to influence the estimate for β.

To address the second issue and get closer to the empirical strategy we used in this chapter, we can estimate the model

$$\log w_{it,t+4} = \beta D_{it} + X_{it}\gamma + \delta_i + \delta_t + \epsilon_{it}.$$

This model still averages the dependent variable, which cannot be undone given the wage data, but uses the baseline value of the democracy index and the controls for the years $1960, 1965, \ldots, 1990$. The bottom panel in Table 21.A1 presents our results using the original wage data (left panel) and updated wage data (right panel). The estimates for β are significantly smaller and never significant. The comparison between the top panel—which uses Rodrik's original specification—and our preferred specification in the bottom suggests that Rodrik's results are, at least in part, driven by a correlation between wages at t and democracy at $t+1, t+2, t+3$ and $t+4$.

Finally, we present estimates of the model

$$\log w_{it,t+4} = \rho \log w_{it-5,t-1} + \beta D_{it} + X_{it}\gamma + \delta_i + \delta_t + \epsilon_{it}, \tag{21.A2}$$

which comes closest to the empirical specification we used throughout the paper. Table 21.A2 has the same structure as Table 21.2 in the paper and presents several estimates of the dynamic panel model in Equation (21.A1). In this case, the lagged dependent variable also controls for the nontrivial autocorrelation patterns induced by averaging the dependent variable. The results confirm that there is no effect of democracy at time t on average wages between t and $t+4$. Only the GMM estimates show large effects that are almost significant at conventional levels. But these estimates are unreliable because they are significantly above the fixed effect models with different imposed values of ρ (and these estimates should bracket them). Moreover, the estimated ρ is too small compared to the fixed effects estimates (it should typically be larger). We believe that this pattern may be caused by the averaging of the dependent variable, which invalidates the moment conditions of GMM estimation.

Table 21.A2 Effects of democratization on the log of manufacturing wages controlling for worker value added, prices and GDP per capita

			GMM			Assuming AR(1) coefficient				
						$\rho=0$	$\rho=0.25$	$\rho=0.5$	$\rho=0.75$	$\rho=1$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Democracy at t	2.89	2.65	15.42	15.91	13.22	3.84	2.65	1.45	0.25	−0.95
	(3.39)	(4.01)	(9.64)	(10.35)	(10.22)	(3.85)	(3.89)	(4.38)	(5.19)	(6.20)
Dep. Var. lagged		0.25★★★	0.20★	0.21★	0.17					
		(0.09)	(0.12)	(0.12)	(0.11)					
Observations	456	384	297	297	298	384	384	384	384	384
Countries	90	86	79	79	79	86	86	86	86	86
Number of moments			40	38	38					
Hansen p-value			0.52	0.44	0.49					
AR2 p-value			0.21	0.21	0.29					
Democracy changes in the sample	47	45	35	35	35	45	45	45	45	45
Long-run effect of democracy	2.89	3.53	19.31	20.01	15.93	3.84	3.53	2.90	1.00	.
p-Value for the long-run effect	0.40	0.50	0.10	0.11	0.19	0.32	0.50	0.74	0.96	.

Dependent variable is log of average wages between t and $t+4$.

Note: OLS estimates (Columns 1–2) include a full set of country and year fixed effects. Arellano and Bond's GMM estimators of the dynamic panel model (Columns 3–4) remove country fixed effects by taking first differences of the data, or by taking forward orthogonal differences (Column 5) and then constructing moment conditions using predetermined lags of the dependent variable and democracy. Columns 4 and 5 use up to the fifth lag of predetermined variables to create moments, restricting the number of moments used. Columns 6–10 impose different values for the autocorrelation coefficient in the tax revenue as a percentage of GDP series, and estimate the effect of democracy including a full set of country and year fixed effects. All models control for the log of GDP per capita, log of worker value added and log of the price level, but these coefficients are not reported to save space. Robust standard errors, adjusted for clustering at the country level, are in parentheses. ★★★: significant at 1%; ★★: significant at 5%; ★: significant at 10%. We do not report long-run effects and their p-values in Column 10 because they are not defined for $\rho=1$.

Rodrik also estimates models using wage data compiled by the Bureau of Labor Statistics for a smaller set of countries. The very small number of democratizations in this sample (only Portugal, South Korea, and Spain) makes these results less reliable. In any case, using Rodrik's original specification, we find that our democracy measure is associated with a 37% increase in wages (standard error = 14.23), but when we estimate the specification in Equation (21.6), including the lagged dependent variable, the effect becomes smaller and no longer significant.

APPENDIX B. RESULTS USING OTHER MEASURES OF DEMOCRACY

In this section we study whether our results are driven by our new measure of democracy. In particular we use Cheibub et al. (2010) Democracy-Dictatorship data (CGV) and Boix-Miller-Rosato's Complete Dataset of Political Regimes, 1800–2007 (BMR). Both datasets are different updates and revisions of the Przeworski et al. (2000) measure. We estimate our basic dynamic panel model using the log of tax revenue as a percentage of GDP, and the Gini coefficient for net and gross income as dependent variable. We only report fixed effects estimates and the Arellano and Bond GMM estimates for each of these variables.

The top panel in Table 21.A3 presents the results using Cheibub et al. (2010) democracy measure; while the bottom panel presents the results using Boix et al. (2012) democracy measure. We find a similar pattern and similar magnitudes, though our GMM estimates on the tax to GDP ratio are less precise and not significant. Again, there is

Table 21.A3 Effects of democratization on the log of tax revenue as a percentage of GDP per capita, and Gini coefficient of net and gross income

	Tax ratio		**Net Gini**		**Gross gini**	
	OLS	**GMM**	**OLS**	**GMM**	**OLS**	**GMM**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Using Cheibub et al. *(2010)* democracy measure						
Democracy lagged	9.48**	11.44	−0.55	−1.45	−1.02	−1.56
	(3.80)	(7.58)	(0.89)	(1.77)	(0.81)	(1.26)
Dep. Var. lagged	0.27***	0.28***	0.32***	0.35***	0.49***	0.77***
	(0.06)	(0.10)	(0.07)	(0.10)	(0.06)	(0.09)
Observations	942	814	537	420	537	420
Countries	128	125	113	100	113	100
Number of moments		81		81		81
Hansen *p*-value		0.17		0.59		0.34
AR2 *p*-value		0.89		0.02		0.45
Democracy changes in the sample	92	82	47	31	47	31
Long-run effect of democracy	12.98	15.82	−0.80	−2.22	−2.01	−6.87

Continued

Table 21.A3 Effects of democratization on the log of tax revenue as a percentage of GDP per capita, and Gini coefficient of net and gross income—cont'd

	Tax ratio		Net Gini		Gross gini	
	OLS	**GMM**	**OLS**	**GMM**	**OLS**	**GMM**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
p-Value for the long-run effect	0.01	0.12	0.53	0.41	0.21	0.24
Using Boix et al. (2012) democracy measure						
Democracy lagged	9.94***	10.57	−0.43	−1.99	−1.23	−2.16
	(3.10)	(9.06)	(0.88)	(1.65)	(0.86)	(1.46)
Dep. Var. lagged	0.27***	0.28***	0.32***	0.35***	0.49***	0.63***
	(0.06)	(0.10)	(0.07)	(0.10)	(0.06)	(0.11)
Observations	944	816	537	420	537	420
Countries	128	125	113	100	113	100
Number of moments		81		81		81
Hansen *p*-value		0.16		0.61		0.42
AR2 *p*-value		0.91		0.02		0.64
Democracy changes in the sample	92	82	47	31	47	31
Long-run effect of democracy	13.61	14.66	−0.63	−3.08	−2.42	−5.81
p-Value for the long-run effect	0.00	0.24	0.62	0.22	0.17	0.17

Note: Odd columns present OLS estimates with a full set of country and year fixed effects. Even columns present Arellano and Bond's GMM estimators of the dynamic panel model which remove country fixed effects by taking first differences of the data and then constructing moment conditions using predetermined lags of the dependent variable and democracy. All models control for the lag of GDP per capita but these coefficients are not reported to save space. Robust standard errors, adjusted for clustering at the country level, are in parentheses. ***: significant at 1%; **: significant at 5%; *: significant at 10%.

an effect on tax revenue as a percentage of GDP, which holds in a more robust way when we focus on specifications in levels that are not reported here to save space. We also continue to find no robust effect on inequality.

Overall, the results are broadly similar using other measures of democracy, though they are more precise and consistent with our preferred measure—as would be expected if our measure removes some of the measurement error present in other indices. This was one of the main goals for its construction.

REFERENCES

Acemoglu, D., 2006. Modeling inefficient institutions. In: Blundell, R., Newey, W., Persson, T. (Eds.), Advances in Economic Theory. Proceedings of 2005 World Congress. Cambridge University Press, New York, NY.

Acemoglu, D., Robinson, J.A., 2000. Why did the west extend the franchise? Q. J. Econ. 115, 1167–1199.

Acemoglu, D., Robinson, J.A., 2001. A theory of political transitions. Am. Econ. Rev. 91, 938–963.
Acemoglu, D., Robinson, J.A., 2006. Economic Origins of Dictatorship and Democracy. Cambridge University Press, New York, NY.
Acemoglu, D., Robinson, J.A., 2008. Persistence of power, elites and institutions. Am. Econ. Rev. 98, 267–291.
Acemoglu, D., Johnson, S., Robinson, J.A., 2001. The colonial origins of comparative development: an empirical investigation. Am. Econ. Rev. 91 (5), 1369–1401.
Acemoglu, D., Johnson, S., Robinson, J.A., 2002. Reversal of fortune: geography and institutions in the making of the modern world income distribution. Q. J. Econ. 118, 1231–1294.
Acemoglu, D., Johnson, S., Robinson, J.A., Yared, P., 2005. From education to democracy? Am. Econ. Rev. 95, 44–49.
Acemoglu, D., Johnson, S., Robinson, J.A., Yared, P., 2008. Income and democracy. Am. Econ. Rev. 98, 808–842.
Acemoglu, D., Johnson, S., Robinson, J.A., Yared, P., 2009. Reevaluating the modernization hypothesis. J. Monet. Econ. 56, 1043–1058.
Acemoglu, D., Ticchi, D., Vindigni, A., 2011. Emergence and persistence of inefficient states. J. Eur. Econ. Assoc. 9 (2), 177–208.
Acemoglu, D., Egorov, G., Sonin, K., 2012. Dynamics and stability of constitutions, coalitions and clubs. Am. Econ. Rev. 102 (4), 1446–1476.
Acemoglu, D., Naidu, S., Restrepo, P., Robinson, J.A., 2013a. Democracy does cause growth, Unpublished.
Acemoglu, D., Robinson, J.A., Santos, R.J., 2013b. The monopoly of violence: evidence from Colombia. J. Eur. Econ. Assoc. 11 (S1), 5–44.
Acemoglu, D., Robinson, J.A., Torvik, R., 2013c. Why Do voters dismantle checks and balances? Rev. Econ. Stud. 80 (3), 845–875.
Acemoglu, D., Reed, T., Robinson, J.A., 2014. Chiefs: elite control of civil society and economic development in sierra Leone. J. Polit. Econ. 122 (2), 319–368.
Aghion, P., Persson, T., Rouzet, D., 2012. Education and military rivalry. Unpublished, http://scholar.harvard.edu/files/aghion/files/education_and_military_rivalry.pdf.
Aidt, T.S., Dallal, B., 2008. Female voting power: the contribution of women's suffrage to the growth of social spending in Western Europe (1869–1960). Public Choice 134 (3–4), 391–417.
Aidt, T.S., Jensen, P.S., 2009a. The taxman tools Up: an event history study of the introduction of the personal income Tax in western Europe, 1815–1941. J. Public Econ. 93, 160–175.
Aidt, T.S., Jensen, P.S., 2009b. Tax structure, size of government, and the extension of the voting franchise in western Europe, 1860–1938. Int. Tax Public Fin. 16 (3), 362–394.
Aidt, T.S., Jensen, P.S., 2011. Workers of the World Unite! Franchise Extensions and the Threat of Revolution in Europe, 1820–1938. Cambridge Working Papers in Economics 1102, Faculty of Economics, University of Cambridge.
Aidt, T.S., Jensen, P.S., 2013. Democratization and the size of government: evidence from the long 19th century. http://ideas.repec.org/p/ces/ceswps/_4132.html.
Aidt, T.S., Dutta, J., Loukoianova, E., 2006. Democracy comes to Europe: franchise expansion and fiscal outcomes 1830–1938. Eur. Econ. Rev. 50 (2), 249–283.
Aidt, T.S., Daunton, M.J., Dutta, J., 2009. The retrenchment hypothesis and the extension of the franchise in England and Wales. Econ. J 120, 990–1020.
Albertus, M., Menaldo, V., 2012. If you're against them you are with us: the effect of expropriation of autocratic survival. Comp. Polit. Stud. 45 (8), 973–1003.
Albertus, M., Menaldo, V., 2014. Gaming democracy: elite dominance during transition and the prospects for redistribution. Br. J. of Polit. Sci. 44 (3), 575–603.
Alesina, A., Angeletos, G.-M., 2005. Fairness and redistribution. Am. Econ. Rev. 95, 960–980.
Alesina, A., Glaeser, E.L., 2004. Fighting Poverty in the US and Europe: A World of Difference. Oxford University Press, New York, NY.
Alesina, A., La Ferrara, E., 2005. Preferences for redistribution in the land of opportunities. J. Public Econ. 89, 897–931.

Alesina, A., Rodrik, D., 1994. Distributive politics and economics growth. Q. J. Econ. 109, 465–490.
Alesina, A., Tabellini, G., 1989. External debt, capital flight and political risk. J. Inter. Econ. 27 (3), 199–220.
Alesina, A., Baqir, R., Easterly, W., 1999. Pubic goods and ethnic divisions. Q. J. Econ. 114, 1243–1284.
Algan, Y., Cahuc, P., Sangnier, M., 2013. Efficient and Inefficient Welfare States. https://sites.google.com/site/pierrecahuc/unpublished_papers.
Alvaredo, F., Atkinson, A.B., Piketty, T., Saez, E., 2010. The World Top Incomes Database. Available from http://topincomes.g-mond.parisschoolofeconomics.eu/.
Alvarez, J., Arellano, M., 2003. The time series and cross-section asymptotics of dynamic panel data estimators. Econometrica 71 (4), 1121–1159.
Anderson, S., François, P., Kotwal, A., 2011. Clientelism in Indian Villages. http://faculty.arts.ubc.ca/fpatrick/documents/clientAERmay1313.pdf.
Ansell, B., 2010. From the Ballot to the Blackboard: The Redistributive Political Economy of Education. Cambridge University Press, New York, NY.
Arbetman-Rabinowitz, M., Kugler, J., Abdollahian, M., Johnson, K., Kang, K., 2011. Replication Data for: Relative Political Capacity Dataset. http://hdl.handle.net/1902.1/16845 Transresearch Consortium.
Arellano, M., Bond, S.R., 1991. Some specification tests for panel data: Monte Carlo evidence and an application to employment equations. Rev. Econ. Stud. 58 (2), 277–298.
Atkinson, A.B., Micklewright, J., 1992. Economic Transformation in Eastern Europe and the Distribution of Income. Cambridge, Cambridge University Press.
Atkinson, A.B., Piketty, T., Saez, E., 2011. Top incomes in the long Run of history. J. Econ. Liter. 49 (1), 3–71.
Austen-Smith, D., Banks, J.S., 1999. Positive Political Theory I. University of Michigan Press, Ann Arbor, MI.
Avelino, G., Brown, D.S., Hunter, W., 2005. The effects of capital mobility, trade openness, and democracy on social spending in Latin America 1980–1999. Am. J. Polit. Sci. 49 (3), 625–641.
Baland, J.-M., Robinson, J.A., 2008. Land and power: theory and evidence from Chile. Am. Econ. Rev. 98, 1737–1765.
Baland, J.-M., Robinson, J.A., 2012. The political value of land: political reform and land prices in Chile. Am. J. Polit. Sci. 56 (3), 601–619.
Bardhan, P., Bowles, S., Wallerstein, M. (Eds.), 2006. Globalization and Egalitarian Redistribution. Princeton University Press, Princeton, NJ.
Bates, R.H., Block, S.A., 2013. Revisiting African agriculture: institutional change and productivity growth. J. Polit. 75 (2), 372–384.
Battaglini, M., Coate, S.T., 2008. A dynamic theory of public spending, taxation and debt. Am. Econ. Rev. 98 (1), 201–236.
Baum, M.A., Lake, D.A., 2001. The invisible hand of democracy: political control and the provision of public services. Comp. Polit. Stud. 34, 587–621.
Baum, M.A., Lake, D.A., 2003. The political economy of growth: democracy and human capital. Am. J. Polit. Sci. 47, 333–347.
Beard, C.A., 1913. An Economic Interpretation of the US Constitution. The Free Press, New York, NY.
Bénabou, R., 2001. Unequal societies: income distribution and the social contract. Am. Econ. Rev. 90, 96–129.
Bénabou, R., 2008. Ideology. J. Eur. Econ. Assoc. 6, 321–352.
Bénabou, R., Ok, E., 2001. Social mobility and the demand for redistribution: the POUM hypothesis. Q. J. Econ. 116, 447–487.
Bénabou, R., Tirole, J., 2006. Belief in a just world and redistributive politics. Q. J. Econ. 121, 699–746.
Berlinski, S., Dewan, T., 2011. The political consequences of franchise extension: evidence from the second reform Act. Q. J. Polit. Sci. 6 (4), 329–376.
Besley, T.F., 2007. Principled Agents: The Political Economy of Good Government. Oxford University Press, Oxford.
Besley, T.F., Kudamatsu, M., 2006. Health and democracy. Am. Econ. Rev. 96, 313–318.
Besley, T.F., Persson, T., 2011. Pillars of Prosperity. Princeton University Press, Princeton, NJ.
Besley, T.F., Persson, T., Sturm, D., 2010. Political competition, policy and growth: theory and evidence from the United States. Rev. Econ. Stud. 77, 1329–1352.

Blaydes, L.M., Kayser, M., 2011. Counting calories: democracy and distribution in the developing world. Inter. Stud. Q. 55, 887–908.
Blundell, R., Bond, S., 2000. GMM Estimation with persistent panel data: an application to production functions. Econ. Rev. 19 (3), 321–340.
Boix, C., Miller, M., Rosato, S., 2012. A complete data Set of political regimes, 1800–2007. Comp. Polit. Stud. 46 (12), 1523–1554.
Brown, D.S., 1999. Reading, writing and regime type: democracy's impact on primary school enrollment. Polit. Res. Q. 52 (4), 681–707.
Brown, D.S., Hunter, W.A., 1999. Democracy and social spending in Latin America, 1980–92. Am. Polit. Sci. Rev. 93 (4), 779–790
Brown, D.S., Hunter, W.A., 2004. Democracy and human capital formation: educational expenditures in Latin America. Comp. Polit. Stud. 37, 842–864.
Burkhart, R.E., 1997. Comparative democracy and income distribution: shape and direction of the causal arrow. J. Polit. 59 (1), 148–164.
Campello, D., 2011. The great gap: inequality and the politics of redistribution in Latin America. In: Blofield, M. (Ed.), The Politics of Redistribution in Less Developed Democracies: Evidence from Brazil, Ecuador, and Venezuela. Penn State Press, University Park, PA.
Carter, M.R., Morrow, J., 2012. Left, Right, Left: Income Dynamics and the Evolving Political Preferences of Forward Looking Bayesian Voters. http://agecon.ucdavis.edu/people/faculty/michael-carter/docs/nopoumECMA.pdf.
Cascio, E., Washington, E., 2012. Valuing the Vote: The Redistribution of Voting Rights and State Funds Following the Voting Rights Act of 1965. National Bureau of Economic Research, Working Paper No. 17776.
Cheibub, J.A., 1998. Political regimes and the extractive capacity of governments: taxation in democracies and dictatorships. World Polit. 50, 349–376.
Cheibub, J.A., Gandhi, J., Vreeland, J.R., 2010. Democracy and dictatorship revisited. Public Choice 143 (1), 67–101.
Daalgard, C.-J., Hansen, H., Larsen, T., 2005. Income skewness, redistribution and growth: A reconciliation, Unpublished, Department of Economics, University of Copenhagen.
Dasgupta, S., 2013. Legalizing the Revolution. Unpublished, Ph.D. Thesis, Columbia University.
De la O, A.L., Rodden, J., 2008. Does religion distract the poor? Income and issue voting around the world. Comp. Polit. Stud. 41 (4), 437–476.
Drazen, A.M., 2000. Political Economy in Macroeconomics. Princeton University Press, Princeton, NJ.
Dunning, T., 2008. Crude Democracy. Cambridge University Press, New York, NY.
Einhorn, R.L., 2006. American Taxation, American Slavery. University of Chicago Press, Chicago, IL.
Elis, R., 2011. Redistribution Under Oligarchy: Trade, Regional Inequality and the Origins of Public Schooling in Argentina, 1862–1912. Unpublished, Ph.D. Dissertation, Department of Political Science, Stanford University.
Ellman, M., Wantchekon, L., 2000. Electoral competition under the threat of political unrest. Q. J. Econ. 115 (2), 499–531.
Engerman, S.L., Sokoloff, K.L., 2005. The evolution of suffrage institutions in the New World. J. Econ. Hist. 65, 891–921.
Engerman, S.L., Sokoloff, K.L., 2011. Economic Development in the Americas since 1500: Endowments and Institutions. Cambridge University Press, New York, NY.
Fernandez, R., Rogerson, R., 1995. On the political economy of education subsidies. Rev. Econ. Stud. 62, 249–262.
Flemming, J.S., Micklewright, J., 2000. Income distribution, economic systems and transition. In: Atkinson, A.B., Bourguignon, F. (Eds.), Handbook of Income Distribution, first ed. In: vol. 1. Elsevier, Cambridge, pp. 843–918, Chapter 14.
Frank, T., 2005. What's the Matter with Kansas? How Conservatives Won the Heart of America. Holt, New York, NY.
Fujiwara, T., 2011. Voting Technology, Political Responsiveness, and Infant Health: Evidence from Brazil. http://www.princeton.edu/∼fujiwara/papers/elecvote_site.pdf.

Gallego, F.A., 2010. Historical origins of schooling: the role of democracy and political decentralization. Rev. Econ. Stat. 92 (2), 228–243.

Gerring, J., Thacker, S., Alfaro, R., 2012. Democracy and human development. J. Polit. 74(1).

Gil, R., Mulligan, C.B., Sala-i-Martin, X., 2004. Do democracies have different public policies than non-democracies? J. Econ. Perspect. 18, 51–74.

Gradstein, M., Justman, M., 1999a. The industrial revolution, political transition and the subsequent decline in inequality in nineteenth century Britain. Explor. Econ. Hist. 36, 109–127.

Gradstein, M., Justman, M., 1999b. The democratization of political elites and the decline in inequality in modern economic growth. In: Breizes, E.S., Temin, P. (Eds.), Elites, Minorities and Economic Growth. Elsevier, Amsterdam.

Gradstein, M., Milanovic, B., 2004. Does liberté=égalité? A survey of the empirical evidence on the links between political democracy and income inequality. J. Econ. Surv. 18 (4), 515–537.

Gradstein, M., Justman, M., Meier, V., 2004. The Political Economy of Education: Implications for Growth and Inequality. MIT Press, Cambridge.

Guttsman, W.L. (Ed.), 1967. A Plea for Democracy: Ad Edited Selection from the 1867 Essays on Reform and Questions for a Reformed Parliament. Macgibbon & Kee, London.

Harding, Robin, Stasavage, David, 2013. What Democracy Does (and Doesn't do) for Basic Services: School Fees, School Inputs, and African Elections. Unpublished, Department of Politics, New York University.

Hassler, J., Storesletten, K., Mora, J.V.R., Zilibotti, F., 2003. The survival of the welfare state. Am. Econ. Rev. 93 (1), 87–112.

Hendrix, C.S., 2010. Measuring state capacity: theoretical and empirical implications for the study of civil conflict. J. Peace Res. 47 (3), 273–285.

Holton, W., 2008. Unruly Americans and the Origins of the Constitution. Palgrave Macmillan, New York, NY.

Holtz-Eakin, D., Newey, W.K., Rosen, H.S., 1988. Estimating vector autoregressions with panel data. Econometrica 56 (6), 1371–1396.

Huber, E., Stephens, J.D., 2012. Democracy and the Left: Social Policy and Inequality in Latin America. University of Chicago Press, Chicago, IL.

Huber, E., Nielsen, F., Pribble, J., Stephens, J.D., 2006. Politics and Inequality in Latin America and the Caribbean. Am. Sociol. Rev. 71 (6), 943–963.

Husted, T.A., Kenny, L., 1997. The effect of the expansion of the voting franchise on the size of government. J. Polit. Econ. 105, 54–82.

Kaufman, R.R., Segura-Ubiergo, A., 2001. Globalization, domestic politics, and social spending in Latin America: a time-series cross-section analysis, 1973–97. World Polit. 53 (4), 553–587.

Krusell, P., Ríos-Rull, J.-V., 1999. On the size of the US government. Am. Econ. Rev. 89 (5), 1156–1181.

Krusell, P., Ríos-Rull, J.-V., Quadrini, V., 1997. Politico-economic equilibrium and growth. J. Econ. Dynam. Control 21 (1), 243–272.

Kudamatsu, M., 2012. Has democratization reduced infant mortality in sub-Saharan Africa? Evidence from micro data. J. Eur. Econ. Assoc. 10 (6), 1294–1317.

Lapp, N.D., 2004. Landing Votes: Representation and Land Reform in Latin America. Palgrave Macmillan, New York, NY.

Larcinese, V., 2011. Enfranchisement and Representation: Italy 1909–1913. Unpublished, Department of Government, London School of Economics, http://personal.lse.ac.uk/LARCINES/enfranchisement_nov11.pdf.

Lee, W., 2003. Is democracy more expropriative than dictatorship? tocquevillian wisdom revisited. J. Devel. Econ. 71, 155–198.

Lee, C.-S., 2005. Income inequality, democracy, and public sector size. Am. Sociol. Rev. 70 (1), 158–181.

Li, H., Squire, L., Zou, H.-f., 1998. Explaining international and intertemporal variations in income inequality. Econ. J. 108 (1), 26–43.

Lindert, P.H., 1994. The rise of social spending, 1880–1930. Explor. Econ. Hist. 31 (1), 1–37.

Lindert, P.H., 2004. Growing Public: Social Spending and Economic Growth since the Eighteenth Century. Cambridge University Press, New York, NY.

Lizzeri, A., Persico, N., 2004. Why Did the elites extend the suffrage? Democracy and the scope of government, with an application to Britain's 'age of reform. Q. J. Econ. 119, 707–765.

Llavador, H., Oxoby, R.J., 2005. Partisan competition, growth, and the franchise. Q. J. Econ. 120, 1155–1192.
Londregan, J.B., 2000. Legislative Institutions and Ideology in Chile. Cambridge University Press, New York, NY.
Lott Jr., J.R., Kenny, L., 1999. How dramatically did women's suffrage change the size and scope of government? J. Polit. Econ. 107 (6), 1163–1198.
Lundahl, M., 1982. The rationale for apartheid. Am. Econ. Rev. 72, 1169–1179.
Martinez-Bravo, M., Padró-i-Miquel, G., Qian, N., Yao, Y., 2012. The effects of democratization on public goods and redistribution: evidence from China. Working Paper.
McDaniel, T., 1991. Autocracy, Modernization and Revolution in Russia and Iran. Princeton University Press, Princeton, NJ.
McGuire, R.A., 2003. To Form a More Perfect Union: A New Economic Interpretations of the United States Constitutions. Oxford University Press, New York, NY.
McGuire, J.W., 2010. Wealth, Health and Democracy in East Asia and Latin America. Cambridge University Press, New York, NY.
McMillan, J., Zoido, P., 2004. How to subvert democracy: montesinos in Peru. J. Econ. Perspect. 4, 69–92.
Meltzer, A.M., Richard, S.F., 1981. A rational theory of the size of government. J. Polit. Econ. 89, 914–927.
Milanovic, B., 1998. Income, Inequality, and Poverty During the Transition from Planned to Market Economy. World Bank, Washington.
Miller, G., 2008. Women's suffrage, political responsiveness, and child survival in American history. Q. J. Econ. 123 (3), 1287–1327.
Moene, K.O., Wallerstein, M., 2001. Inequality, social insurance and redistribution. Am. Polit. Sci. Rev. 95, 859–874.
Mohammed, S., Finnoff, K., 2003. Capital Flight from South Africa, 1980–2000. In: Epstein, G. (Ed.), Capital Flight and Capital Controls in Developing Countries. Edward Elgar Publishing, Inc, Northampton.
Moore, B., 1966. The Social Origins of Dictatorship and Democracy: Lord and Peasant in the Making of the Modern World. Beacon Press, Boston, MA.
Moses, J.W., 1994. Abdication from national policy autonomy: what's left to leave? Polit. Soc. 22 (2), 125–148.
Muller, E.N., 1988. Democracy, economic development, and income inequality. Am. Sociol. Rev. 53 (1), 50–68.
Muller, E.N., 1989. Democracy and inequality. Am. Sociol. Rev. 54 (5), 868–871.
Naidu, S., 2011. Suffrage, Schooling, and Sorting in the Post-Bellum US South. http://tuvalu.santafe.edu/∼snaidu/papers/suffrage_may_25_2012_combined.pdf.
Naidu, S., Yuchtman, N., 2013. Coercive contract enforcement: law and the labor market in 19th century industrial Britain. Am. Econ. Rev. 103 (1), 107–144.
Newey, W.K., Windmeijer, F., 2009. GMM estimation with many weak moment conditions. Econometrica 77 (3), 687–719.
Nickell, S., 1981. Biases in dynamic models with fixed effects. Econometrica 49 (6), 1417–1426.
Panitch, L., Gindin, S., 2012. The Making of Global Capitalism. Verso, New York, NY.
Papaioannou, E., Siourounis, G., 2008. Democratisation and growth. Econ. J. 118 (532), 1520–1551.
Perotti, R., 1996. Growth, income distribution and democracy: what the data say. J. Econ. Growth 1, 149–187.
Persson, T., Tabellini, G., 1994. Is inequality harmful to growth? Am. Econ. Rev. 84, 600–621.
Persson, T., Tabellini, G., 2000. Political Economics. MIT Press, Cambridge.
Persson, T., Tabellini, G., 2003. The Economic Effects of Constitutions. MIT Press, Cambridge.
Pettersson-Lidbom, P., Tyrefors, B., 2011. Democracy, Redistribution, and Political Participation: Evidence from Sweden 1919–1938. http://people.su.se/∼pepet/democracyFeb2012.pdf.
Piketty, T., 1995. Social mobility and redistributive politics. Q. J. Econ. 110, 551–584.
Przeworski, A., Alvarez, M., Cheibub, J.A., Limongi, F., 2000. Democracy and Development: Political Institutions and Material Well-being in the World, 1950–1990. Cambridge University Press, New York, NY.

Putterman, L., 1996. Why have the rabble not redistributed the wealth? On the stability of democracy and unequal wealth. In: Roemer, J.E. (Ed.), Property Relations, Incentives and Welfare. McMillan, London.

Roberts, K.W.S., 1977. Voting over income tax schedules. J. Public Econ. 8, 329–340.

Rodrik, D., 1999. Democracies pay higher wages. Q. J. Econ. 114, 707–738.

Roemer, J.E., 1998. Why the poor do not expropriate the rich: an old agrument in new garb. J. Public Econ. 70, 399–424.

Roemer, J.E., Lee, W., Van der Straeten, K., 2007. Racism, Xenophobia and Distribution: Multi-Issue Politics in Advanced Democracies. Harvard University Press, Cambridge.

Romer, T., 1975. Individual welfare, majority voting and the properties of a linear income tax. J. Public Econ. 7, 163–168.

Ross, M.L., 2006. Is democracy good for the poor? Am. J. Polit. Sci. 50, 860–874.

Scheve, K., Stasavage, D., 2009. Institutions, Partisanship, and Inequality in the Long Run. World Polit. 61, 215–253.

Scheve, K., Stasavage, D., 2010. The conscription of wealth: mass warfare and the demand for progressive taxation. Inter. Organ. 64, 529–561.

Scheve, K., Stasavage, D., 2012. Democracy, war, and wealth: lessons from two centuries of inheritance taxation. Am. Polit. Sci. Rev. 106 (1), 82–102.

Schonhardt-Bailey, C., 2006. From the Corn Laws to Free Trade: Interests, Ideas, and Institutions in Historical Perspective. MIT Press, Cambridge.

Seligmann, L.J., 1995. Between Reform & Revolution: Political Struggles in the Peruvian Andes, 1969–1991. Stanford University Press, Stanford, CA.

Shayo, M., 2009. A model of social identity with an application to political economy: nation, class, and redistribution. Am. Polit. Sci. Rev. 103 (02), 147–174.

Siavelis, P.M., 2000. The President and Congress in Post-Authoritarian Chile: Institutional Constraints to Democratic Consolidation. Pennsylvania State University Press, University Park, PA.

Simpson, M., 1990. Political rights and income inequality: a cross-national test. Am. Sociol. Rev. 55 (5), 682–693.

Sirowy, L., Inkeles, A., 1990. The effects of democracy on economic growth and inequality: a review. Stud. Comput. Inter. Develop. 25, 126–157.

Soifer, H., 2013. State power and the redistributive threat. Stud. in Comp. Int. Dev. 48 (1), 1–22.

Solt, F., 2009. Standardizing the world income inequality database. Soc. Sci. Q. 90 (2), 231–242.

Stasavage, D., 2005a. Democracy and education spending in Africa. Am. J. Polit. Sci. 49, 343–358.

Stasavage, D., 2005b. The role of democracy in Uganda's move to universal primary education. J. Mod. Afr. Stud. 43 (1), 53–73.

Stigler, G.J., 1970. Director's law of public income redistribution. J. Law Econ. 13, 1–10.

Tilly, C., 1985. Warmaking and state making as organized crime. In: Evans, P., Rueschemeyer, D., Skocpol, T. (Eds.), Bringing the State Back In. Cambridge University Press, Cambridge.

Vanhanen, T., 2013. Democratization and Power Resources 1850–2000. http://www.fsd.uta.fi/en/data/catalogue/FSD1216/meF1216e.html#viittaaminen.

Waldstreicher, D., 2009. Slavery's Constitution: From Revolution to Ratification. MacMillan, New York, NY.

Weede, E., 1989. Democracy and income inequality reconsidered. Am. Sociol. Rev. 54 (5), 865–868.

Weyland, K., 2004. Neoliberalism and democracy in Latin America: a mixed record. Lat. Am. Polit. Soc. 46 (1), 135–157.

Whiteford, A., Van Seventer, D., 2000. Understanding contemporary household inequality in South Africa. Stud. Econ. Econ. 24 (3), 7–30.

Wigley, S., Akkoyunlu-Wigley, A., 2011. The impact of regime type on health: does redistribution explain everything? World Polit. 63 (4), 647–677.

Wilse-Samson, L., 2013. Structural Change and Democratization: Evidence from Apartheid South Africa. Unpublished working paper, Columbia University, http://www.columbia.edu/~lhw2110/research.html.

Wooldridge, J.M., 2002. Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge.

Wright, R.D., 1996. Taxes, redistribution and growth. J. Public Econ. 62, 327–338.

CHAPTER 22

The Idea of Antipoverty Policy

Martin Ravallion
Department of Economics, Georgetown University, Washington, DC, USA

Contents

Handbook of Income Distribution, Volume 2B
ISSN 1574-0056, http://dx.doi.org/10.1016/B978-0-444-59429-7.00023-6

Abstract

How did we come to think that eliminating poverty is a legitimate goal for public policy? What policies emerged in the hope of attaining that goal? The last 200 years have witnessed a dramatic change in thinking about poverty. Mainstream economic thinking in the eighteenth century held that poverty was necessary and even desirable for a country's economic success. Today, poverty is more often viewed as a constraint on that success. In short, poverty switched from being seen as a social good to a social bad. This change in thinking, and the accompanying progress in knowledge, has greatly influenced public action, with heightened emphasis on the role of antipoverty policy in sustainable promotion from poverty, as well as protection. Development strategies today typically strive for a virtuous cycle of growth with equity and a range of policy interventions have emerged to help ensure that outcome. An expanding body of knowledge has taught us about how effective those interventions are in specific settings, although many knowledge gaps remain.

Keywords

Poverty, Inequality, Growth, Redistribution, Antipoverty policy

JEL Classification Codes

B00, I38, O15

22.1. INTRODUCTION

The poor . . . are like the shadows in a painting: they provide the necessary contrast.

Philippe Hecquet (1740), quoted in Roche (1987, p. 64)

Everyone but an idiot knows that the lower classes must be kept poor or they will never be industrious.

Arthur Young (1771), quoted in Furniss (1920, p. 118)

May we not outgrow the belief that poverty is necessary?

Marshall (1890, p. 2)

Our dream is a world free of poverty.

(Motto of the World Bank since 1990).

It is widely accepted today that eliminating poverty is a legitimate goal of public action, for which governments (in both rich and poor countries) typically take some responsibility. The policy responses include both direct interventions, often put under the heading of "social policies," and various economy-wide policies—overall policies for economic development that have bearing on the extent of poverty. (I will use the term "antipoverty policy" to embrace both sets of policies.) There are essentially three premises to the idea of such policies:

- Premise 1: Poverty is a social bad.[1]
- Premise 2: Poverty can be eliminated.
- Premise 3: Public policies can help do that.

This chapter tries to understand how these three premises came to be broadly accepted and what forms of public action emerged.

Both the differences and the similarities between today's thinking and that of the past are of interest. There are some policy debates that live on and some common themes, such as the role of incentives. However, one is also struck by the differences. Indeed, widespread (though certainly not universal) acceptance of the three premises above appears to be relatively new. Before the late eighteenth century, the dominant school of economic thought saw poverty as a social *good*, essential for economic development. It may well have been granted that, other things being equal, a society with less poverty is to be preferred, but other things were not seen to be equal. Poverty was deemed essential to incentivize workers and keep their wages low, so as to create a strong, globally competitive, economy. Nor did the idea of what constitutes "economic development" embrace poor people as being necessarily among its intended beneficiaries. There was also widespread doubt about the desirability of, or even the potential for, governmental intervention against poverty. In marked contrast, poverty is widely seen today as a constraint on development rather than a precondition for it. And it is now widely (though not universally) agreed, across both rich and poor countries, that the government has an important role in the fight against poverty.

This chapter documents this transition in thinking about poverty and policy. Of course, the interrelationship between thinking and action is complex, and what emerges in the policy arena depends on many things, including technology, public awareness, and political economy. Nonetheless, there is a story to be told about how scholarly and popular thinking has evolved. This helps us understand prevailing views on the distributive role of the state and the specific policies adopted. The change in thinking also teaches us that the progress in knowledge both reinforces and reflects progress in development.

A natural starting point is Fleischacker's (2004) excellent *Short History of Distributive Justice*. Fleischacker defines distributive justice as a situation in which "property is distributed throughout the society so that everyone is supplied with a certain level of material means."[2] He argues that in premodern thought, poverty relief was largely motivated by

[1] Poverty can be seen as a social bad either intrinsically (a society with less poverty is preferred) or instrumentally (that a less poor society will be better at other things of value, including its overall economic performance).

[2] Aristotle is widely credited with introducing the term "distributive justice," in the fourth century BC. However, Fleischacker (2004) convincingly argues that Aristotle had something quite different in mind to modern usage. For Aristotle, distributive justice was about assigning political rewards according to "merit."

beneficence—a matter of the donor's personal choice, not a right for poor people, and so quite distinct from justice, which emanates from the secular world of laws and taxes. Most religions see voluntary efforts to help poor people as a virtue.[3] However, such charitable relief is not distributive justice in Fleischacker's eyes. For the birth of that idea, he argues that we need to look to Europe in the late eighteenth century. Fleischacker describes and interprets the development of the idea in philosophical writings. However, what is largely missing from Fleischacker's history is the economics. This is important if we focus on poverty rather than justice. Nor have historians (such as Beaudoin, 2007; Geremek, 1994; Himmelfarb, 1984a,b) given more than passing attention to the economics. And it would be fair to say that economists have paid little attention to the history of thought on poverty and inequality.[4]

The chapter offers an overview of how philosophical and economic thinking on poverty and antipoverty policy has evolved and the types of policies that emerged. The discussion will give less emphasis than Fleischacker on whether poor people were believed to have the legal right to assistance. States can and do ascribe legal rights, but sometimes with little more than symbolic value, given that the administrative capabilities for enforcement are weak, and especially so in poor countries. Instead, the focus here will be on whether (demonstrably or plausibly) public policy helped families permanently escape poverty or merely offered a transient (though potentially important) short-term palliative to protect people from negative shocks. In short, the acid test for a good antipoverty policy will be whether it is aimed at both *promotion* and *protection* (applying a useful distinction made by Drèze and Sen, 1989). This idea of antipoverty policy turns out to be quite recent, with origins in the late nineteenth century, but only emerging with confidence in the late twentieth century.

The chapter begins with a simple characterization of personal wealth dynamics, which will help motivate the chapter's interpretation of past thinking about this class of policies. The bulk of the chapter falls into two parts. The first, comprising Sections 22.3–22.6, traces out the history of thought from mercantilist views on the inevitability of poverty through two main stages of "poverty enlightenment," out of which poverty came to be seen as a social bad (Premise 1). The second part focuses on policies, from economy-wide policies to direct interventions. Sections 22.7 and 22.8 turn to an important aspect of Premise 2, namely, the country's overall development strategy and in particular whether poverty can be eliminated through economic growth and the role played by initial

[3] Although with differences in emphasis, both within and between religions. For example, see Kahl's (2005) discussion of the differences between Catholicism, Lutheranism, and Calvinism/Reformed Protestantism.

[4] In one of the few exceptions, Cogneau (2012) discusses the evolution of thought on inequality in a development context. On the neglect of the history of thought by economists, see Blaug (2001).

distribution. Section 22.9 focuses on present-day thinking on specific direct interventions (Premise 3). Section 22.10 concludes the chapter.

22.2. WEALTH DYNAMICS AND ANTIPOVERTY POLICIES

A longstanding explanation heard for poverty is that it stems from the "bad behaviors" of poor people—high fertility, laziness, or bad spending choices, such as excessive consumption of alcohol.[5] It is not that they are in any way constrained to be poor, but that they (implicitly or explicitly) chose to be poor. By this view, the role for antipoverty policy is to ensure behavioral change. We will hear more of these arguments later in this chapter. However, it will be useful to sketch here an alternative model whereby poverty emerges from the wealth dynamics implied by the external constraints facing poor people.

By "wealth" I shall mean both human capital—the accumulated stock of past educational and health inputs, including past nutritional intakes—as well as nonhuman capital, such as industrial or financial capital.[6] To simplify the analysis, however, wealth is treated as a single composite asset. Initial wealth, w_t at date t, is distributed across individuals, some of whom have zero wealth, but may still earn some labor income, consumed fully on their survival needs in each period. A fixed share of current wealth is used for current consumption. Each person has a production function yielding output $h(k)$ from a capital stock k. There is a threshold capital stock needed to produce any output, i.e., $h(k)=0$ for all $k \leq k^{\min}(>0)$. Once the threshold is reached, output emerges in the next period, though diminishing returns start to set in immediately; in other words, the function $h(k)$ is strictly positive, strictly increasing and strictly concave for all $k>k^{\min}$. Those for whom the threshold has not been reached ($w<k^{\min}$) have no demand for capital as it will not yield any output.

There is more than one interpretation of the threshold. Dasgupta (1993) provides a persuasive argument for its existence based on the biological fact of a positive basal metabolic rate, given that maintaining the human body at rest requires a (substantial) minimum food energy intake, without which no physical work can be done. (Maintenance requirements are 60–75% of food energy intake.) Physiology entails that the set of feasible production activities for an individual is inherently nonconvex. Threshold effects can also reflect nonconvexities in production possibilities associated with minimum schooling needs, the nature of the production technology or from the existence of a lumpy "threshold good" in consumption.[7] In a more elaborate version of this model, one would

[5] See Klebaner (1964), Burnett (1969, Chapter 4), and Wim and Halman (2000).

[6] A good typology is found in Sachs's (2005b, Chapter 13) six types of capital that poor people lack: human capital, business capital, infrastructure, natural capital, public institutional capital, and knowledge capital.

[7] On the latter argument, see Just and Michelson (2007). On other sources of poverty traps, see Azariadis (2006) and other papers in the collection edited by Bowles et al. (2006).

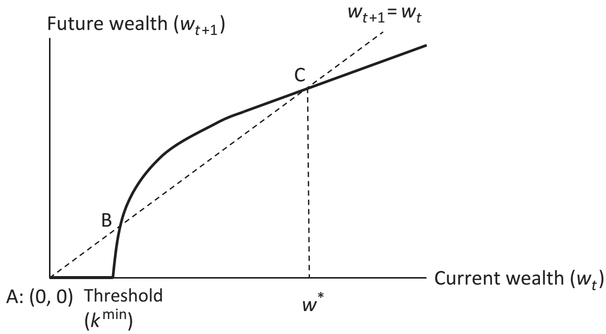


Figure 22.1 Wealth dynamics with a poverty trap.

also want to allow for interaction effects among different dimensions of wealth, such as when poor nutritional status impedes children's learning.

There is another constraint on production possibilities stemming from credit market failures. Because lenders are imperfectly informed about borrowers, a borrowing constraint is imposed, whereby a person can only borrow up to λ times her wealth. Let $k^\star$ denote the individual's desired capital stock. Those with wealth sufficient to produce but less than $k^\star/(\lambda+1)$ have a desire to invest but are constrained in that, after investing all they can, they still find that the marginal product of capital exceeds the interest rate, given the borrowing constraint. Finally, someone who starts her productive life with sufficient wealth (greater than $k^\star/(\lambda+1)$) is able to invest her unconstrained optimal amount, equating the (declining) marginal product of her capital with the prevailing interest rate r (the price of capital), which is taken to be fixed ($h'(k^\star)=r$).[8]

The recursion diagram (the mapping from current wealth to future wealth) then takes the form depicted in Figure 22.1. Future wealth is zero at low levels of current wealth ($w_t < k^{\min}$). For levels of initial wealth in the interval $[k^{\min}, k^\star/(\lambda+1)]$, future wealth is a strictly concave function of current wealth. At higher wealth ($w_t > k^\star/(\lambda+1)$), the function becomes linear.

There are potentially three steady-state equilibria (with constant wealth over time) for each individual. Two of these, namely, points A and C in Figure 22.1, are stable while the middle one, at point B, is unstable in that shocks will move those at B toward A or C.[9] In the long run, after repeated small shocks, the economy will settle in a state that can be thought of as having two main classes of people. One class has little or no wealth, given that its members are caught in a *wealth poverty trap*, at point A. There can be many reasons

[8] In the special case in which the threshold is not binding, this model is the same as that outlined in Banerjee and Duflo (2003), though with antecedents in the literature.

[9] Imagine someone at point B in Figure 22.1. Any small wealth gain will put her in a region of accumulation (current wealth lower than future wealth) and so the person will progress toward point C. Similarly, a small contraction will put her on a path to point A.

in practice why people are so trapped, including lack of any marketable skills, social exclusion, geographic isolation, debilitating disease, or environmental degradation. The second class comprises people who have settled at point C, at their respective steady-state levels of wealth (w^*). There can still be inequality within each class. There can be inequality of labor earnings among the poorer class, and there can be wealth inequality among the "point C folk," given different steady-state levels of wealth. There can be poverty even if nobody is caught in a poverty trap. The "poor" can be identified as two groups of people, namely, those at point A and the poor among those at point C, that is, those for whom their steady-state level of wealth turns out to be very low, even though they are not caught in a poverty trap.

Although the wealth poverty trap at point A is economically stable for each individual, social and political stability is another matter. The latter types of instability can arise in many ways, defying simple generalizations about its economic causes. However, it is plausible that a large mass of people at point A can threaten social stability, especially if their labor earnings and (hence) consumptions are very low, either in steady state or as a result of some severe shock, and in the latter case the threat to stability may well be even greater.[10]

Motivated by this stylized representation of wealth dynamics, we can think of two broad types of antipoverty policies. There can be policies that provide short-term palliatives, possibly to maintain social stability by assuring that current incomes do not fall below some crucial level, even though poor people remain poor, either because they are caught in a wealth poverty trap or they have a low steady-state level of wealth. These are purely *protection policies*. And there are *promotion policies* that allow poor people to attain the higher level of wealth needed to escape poverty. For those caught in a poverty trap, this will require a sufficiently large wealth gain to put them on a path to eventually reaching their own (higher and stable) steady-state level of wealth. For those not caught in a trap, but still poor, promotion will require some combination of higher wealth and higher returns to their wealth—an upward shift in the recursion diagram in Figure 22.1.

The rest of this chapter will study the origins and nature of both types of policies. It will be argued that, although the idea of some public responsibility for protecting poor people from negative shocks is an old one, the idea of such a role for promotion by relieving the constraints facing poor people (either caught in poverty trap or with low returns to their wealth) is remarkably new. The latter idea came with a significant evolution in thinking about the causes of poverty. The longstanding view that the "moral weaknesses" of poor people caused their poverty implied little scope for public action to promote people from poverty, and rebuffed any calls for taxing the rich to finance such action. It was ultimately up to poor people to escape poverty by changing their behaviors. Public responsibility was largely confined to limited, and highly targeted, protection to address extreme transient poverty and some

[10] In *Politics*, Aristotle (350 BC, unnumbered) put the point nicely: "It is a bad thing that many from being rich should become poor; for men of ruined fortunes are sure to stir up revolutions."

efforts at aiding the "moral reform" of poor people. Although one still hears casual claims blaming poor people for their poverty today, across the globe, from the mid-nineteenth century, though not carrying much policy weight until well into the twentieth century, deeper causal understandings of poverty emerged in popular and scholarly writings. These pointed to a new promotional role for public action in fighting persistent poverty.[11] Poverty was seen to reflect in no small measure public failures, including uncorrected market failures.

22.3. THE UTILITY OF POVERTY

For much of the sixteenth through eighteenth centuries, when Western Europe was mired in poverty, the dominant economic theory of the time, mercantilism, saw poverty as a natural state of affairs and, indeed, instrumentally good, as a means of encouraging work effort. The mercantilist goal was to maximize a nation's export surplus—the balance of trade, which was equated with the future prosperity and power of the realm—and the means were cheap production inputs, that is, cheap raw materials (for which colonies proved useful) and cheap, and therefore poor, labor at home. Poverty was not just accepted; it was seen as an essential precondition for a country's economic development. Hunger would encourage work, and lack of it would do the opposite. The seemingly widely held economic premise was that the individual supply curve for unskilled work was negatively sloped—in modern economic terms, that the income effect on demand for leisure dominated the substitution effect. As the Reverend Townsend (1786) put it: "The poor know little of the motives which stimulate the higher ranks to action—pride, honor and ambition. In general, it is only hunger which can spur and goad them onto labor" (p. 23). And so: "... in proportion as you advance the wages of the poor, you diminish the quantity of their work" (p. 29).[12]

The idea of a negatively sloped labor supply curve is essentially what Furniss (1920, p. 117) later dubbed "the utility of poverty." The basis for this idea appears to have been little more than casual anecdotes; Furniss (1920, Chapter 6) provides many examples from writings of the time, often with references to the attractions of the alehouse when workers got a wage increase. It was not the last time in the history of thought about poverty that casual incentive arguments resting on little or no good evidence would buttress strong policy positions.

A continuing future supply of cheap labor was also seen to be crucial. Large families were encouraged and good work habits were to be instilled from an early age. Like higher current wages, too much schooling would discourage both current and future work

[11] This is not to say that the change in the model of poverty caused the change in policy. To some extent, both changes shared a common causation in broader changes in the economy and society.

[12] Though little known today, Townsend's advocacy of free markets was important in the history of economic thought, with influence on subsequent thinkers (including Malthus and Darwin). For further discussion of Townsend's influence, see Montagu (1971) and Lepenies (2014).

effort. Consistent with this model, few sustainable opportunities were expected to be available to any educated children from poor families. In de Mandeville's (1732, pp. 288–311) mind, the only realistic future prospect for the children of laboring (and hence poor) parents was to be laboring and poor. Poor parents had little realistic hope that their children would be anything but poor; their low aspirations simply reflected and rationalized their lack of opportunity. Small amounts of schooling would have served little purpose. In this view of economic development, there was little or no prospect of reducing wealth poverty—including escaping the poverty trap demonstrated in Figure 22.1. There was little or no perceived scope for upward mobility of working class children. They were born poor and stayed poor.

Modern progressive thinkers may be shocked by de Mandeville's views (and similar views still heard occasionally in modern times), but there may well be an element of cruel truth to them. His claim that a modest amount of extra schooling for working class children is wasted is consistent with the model in Figure 22.1. Suppose that the poor—the working class—are concentrated at the wealth poverty trap (point A in Figure 22.1). A small increase in their wealth, in the form of extra human capital only sufficient to get them to the threshold (say), will not bring any lasting benefit. In due course, the dynamics will push them back to point A. A large gain in schooling is needed.

And de Mandeville's pessimism about schooling would not surprise many poor children in the developing world today. Katherine Boo's (2012) vivid description of life in a Mumbai slum includes a discussion of the choices made by Sunil, a young scavenger who spends long hours collecting whatever he can find of any value in the trash deposits around Mumbai airport. Sunil is clearly very poor. He is also clearly capable of learning and is aware that with sufficient schooling he might escape his wretched life. But how can he finance sufficient schooling? At one point, he spends a few days in a private after-hours school run by a college student who lives in the slum, and after much rote learning he masters the "twinkle-star" song.[13] Boo (2012, p. 68) writes:

> *He'd sat in on [the English class taught in the slum] for a few days, mastering the English twinkle-star song, before deciding that his time was better spent working for food.*

By interpretation, the modest amount of schooling that Sunil could afford would be insufficient for him to escape poverty. He is better off addressing his current hunger.

22.3.1 Early Social Protection Policies

Recall that the poverty trap in Figure 22.1 has people stuck at zero wealth but they still earn enough to survive (as Sunil does through scavenging). Higher wages or prices for their outputs increase their welfare, and uninsured shocks to their health (say) have the

[13] "Twinkle-twinkle little star, how I wonder what you are. Up above the world so high, like a diamond in the sky."

opposite effect. There is space here for social protection policies providing state-contingent income support. Such policies can exist and be seen as reasonably effective without changing the fact that poor people are stuck in the wealth poverty trap.

It has long been argued that governments have a role in social protection from shocks that threaten extreme poverty. For example, around 300 BC, the famous Indian academic and advisor to royalty, Chanakya (also known as Kautilya), recommended that when famine looms a good king should ". . . institute the building of forts or water-works with the grant of food, or share [his] provisions [with the people], or entrust the country [to another king]" (quoted in Drèze, 1990a, p. 75). If one thinks of antipoverty policy primarily in terms of protection from adverse events, then the idea is very old indeed.

Even though mainstream economic thinking has for a long time encouraged a limited role for the state in social protection, more contentious has been the idea of promotion. In the premercantilist feudal and slave economies, the employer had a responsibility for insuring workers, even very poor workers, who may well have faced exploitation but were at least protected to some degree. (This was not necessarily altruistic in any sense; a slave owner had a purely selfish interest in keeping his property alive.) The new elites in the early development of capitalism were keen to see the state take over these roles, but consistently with their economic ideas. The status quo distribution of wealth was seen by its defenders as the outcome of natural processes, which included the competitive market mechanism, and it was not to be tampered with through policy. Persistent poverty was believed to be the natural order of things until modern times. By contrast, transient poverty was seen as a threat to the social order. There was at least an implicit recognition of the limitations of free markets in providing insurance against risk.

The sixteenth and seventeenth centuries saw the emergence of fledgling social policies in Europe in response to rising "pauperism." There were increasing numbers of dislocated and unemployed workers and beggars on city streets. Although the cause was widely seen to be the moral weaknesses of poor people, deeper explanations could be found in changes in the organization of production (including in agriculture with the breakup of feudalism) combined with greater mobility (also with implications for family support of the aged). Although unemployment was not commonly identified as a cause of poverty, work was widely seen as the solution. Publicly financed workhouses were introduced around 1600. Welfare recipients were incarcerated and obliged to work for their upkeep. From the outset, the idea was that the workhouses would be "self-targeting," in that only the poorest would be willing to be so confined, thus providing a cost-effective means of poverty relief (Thane, 2000, p. 115). But the policy was also grounded in the prevailing view that poverty was caused by bad behaviors, which could be controlled and (hopefully) corrected by the workhouses. The workhouses were seen as a cost-effective policy for moral reform.

There was a strong element of protection in the workhouse idea; anyone thrown into poverty by some shock could turn to the workhouse. Was it also a promotional policy?

There is not much discussion in the literature about the promotional value to poor people of the work done beyond the perceived moral value of actually doing work. Advocates might well argue that this was promotion through behavioral change. But it was clearly not promotion by relieving the constraints facing poor people.

22.3.2 England's Poor Laws

A major policy response to poverty emerged in Elizabethan England in the form of the Poor Laws.[14] This was a system of publicly provided insurance against income poverty due to specific sources, notably old age, widowhood, disability, illness, or unemployment. Essentially the central government instructed local parishes to deal with their poverty problem. As a system of protection, the Poor Laws were quite comprehensive and came to be reasonably generous in some places.[15] Arguably the pinnacle of the Poor Laws was the Speenhamland System of 1795 introduced by the justices of Berkshire. This system aimed to ensure a guaranteed minimum income through a sliding scale of wage supplements indexed to the price of bread (Himmelfarb, 1984a; Montagu, 1971).

The antipoverty programs elsewhere in Europe around this time relied heavily on charitable giving and so faced free-rider problems; levels of church and private spending on transfers to poor people were low—well under 1% of national income in most countries (Lindert, 2013). In contrast, the disbursements under the Poor Laws in England and Wales were largely financed by local property taxation. There was evidently some displacement of private charity, though the latter continued to exist (Hindle, 2004; Lindert, 2013). But there can be little doubt that the Poor Laws entailed a net gain in social protection. By the late seventeenth century almost all parishes of England and Wales were covered, and, under the "Old Poor Laws" up to the nineteenth century, all persons were eligible for relief. (New Poor Laws came out of reforms in the 1830s, which I return to later.) The parishes had the responsibility for implementation, subject to monitoring by central authorities. Being based in the parishes was convenient but possibly never ideal as they provided limited scope for pooling risks, and there was undoubtedly considerable horizontal inequity (whereby equally poor people in different parishes fared very differently).[16] Nor could these policies ever be expected to have much impact on the steady-state wealth distribution. However, it is clear that the

[14] On the history of the English Poor Laws and their influence, see Mencher (1967), Boyer (2002), and Hindle (2004).

[15] Solar (1995) cites evidence that aggregate disbursements reached 2% of England's national income by the late eighteenth century.

[16] Hindle (2004) notes the large geographic differences in pensions, depending on the economic circumstances of the parishes.

Old Poor Laws did provide a degree of protection from risk, and it has been argued that they helped break the historical link between harvest failures and mortality (Kelly and Cormac, 2010; Smith, 2011).

The Poor Laws appear to have helped ensure a relatively docile and sustained working class, with little threat to the steady-state distribution of wealth. Solar (1995) argues that the Old Poor Laws were crucial to England's long-term social stability, including periods (such as the late eighteenth century) of concern about the possibility of the dramatic instability in France spilling across the English Channel. Broad political support was ensured by the fact that anyone could get relief if needed. For example, widowhood was a threat to many of those who would not normally expect to turn to the parish for relief.[17] As novels of the time often pointed out, even the well-to-do upper middle class family could be vulnerable to poverty (a favorite theme of Charles Dickens).

Fleischacker's (2004, p. 51) discussion of England's Poor Laws argues that they were motivated by the "… virtue of charity rather than the virtue of justice," and as such they did not constitute the beginnings of the modern role for public policy in assuring distributive justice. One can conjecture that the motivation for the Poor Laws was at least as much to do with maintaining social stability as charity or justice. However, whatever may have been the motives of policy makers, the Poor Laws constituted a legally enforceable state policy for limited relief from the specified events, financed by redistributive taxes. And parish residents (though not outsiders) had a legal recourse under the Poor Laws, which is why they could help ensure social stability over some 300 years (Solar, 1995). Against Fleischacker's interpretation, it seems that the Poor Laws came very close to being a premodern example of policies to help ensure distributive justice.

However, an aspect of the Poor Laws that should not be ignored is that they were clearly intended for protection rather than promotion. These laws were an early form of social insurance intended for a world in which the poor and the middle class faced many uninsured risks associated with uncertain employment, health crises, harvest failures, and simple bad luck (Hindle, 2004). Such risks may well have spilled over into production, with adverse long-term consequences. By assuring greater social stability, this, too, may have brought long-term gains. However, it is clear that any longer-term promotional advantages were attained via the enhanced protection that was attained under the Poor Laws. Protection was clearly seen as the main aim of the Poor Laws.

[17] Widows were listed as eligible for relief from the earliest Poor Laws, and they are mentioned often in the literature; for example, Hindle's (2004) discussion of parish archival information related to the Poor Laws mentions widows 75 times.

Instead of focusing on whether the motivation was charity or justice, the more important reason why the Elizabethan Poor Laws, or Chanakya's famine relief policy, did not constitute a comprehensive antipoverty policy is that these policies were unlikely to change the steady-state distribution of the levels of wealth. In terms of the model in Section 22.2, what these policies were doing was preventing the consumption levels of those either stuck in the wealth poverty trap or settled at some low steady-state level of wealth from falling too much. They provided a degree of protection but did little to help people permanently escape poverty. By the economic logic of the mercantilists, hunger was a good thing as it motivated poor people to work, with social protection playing a limited and well-defined role. After all, just like the slave owner, mercantilists believed that one must keep the workers alive.

By the late eighteenth century, a significant change in thinking was underway.

22.4. THE FIRST POVERTY ENLIGHTENMENT

The incidence of poverty had clearly been increasing for some time in Britain and much of Europe in the latter part of the eighteenth century, due mainly to falling real wages (Allen, 2007; Tucker, 1975). In Europe and North America, there was mounting concern about prospects for social instability and even rebellion among the working class. There was also frustration among the middle class about the constraints they faced on their upward mobility. And there were clearly some gaping weaknesses in the prevailing mainstream intellectual defenses of the status quo. Inherited inequalities of opportunity and manipulated noncompetitive market processes (sometimes facilitated by government) started to be seen as playing an important role in determining the distribution of wealth, casting doubt on claims that the status quo distribution was some purely natural order emerging from free markets.

The masses started to question longstanding excuses for the deprivations they faced. Of course, there had been sporadic propoor protest movements before. For example, there was the (short-lived) "Levellers" movement for suffrage and religious tolerance in mid-seventeenth-century England, during the Civil War period (Hill, 1972). But the late eighteenth century saw both new thinking and more widespread demands for change across Britain, Europe, and America. Popular politics flourished in the cafés and alehouses of London, Paris, and elsewhere in Europe in the late eighteenth century.[18] The historian Brinton (1934, p. 281) identifies the "essential characteristic" of the change in ideas in the last decade of the eighteenth century in Europe as the transition from the view that ". . . life on this earth is a fleeting transition to eternity, that such life is inevitably

[18] The Proceedings of the Old Bailey (2012) contains descriptions for London; an example was the "London Corresponding Society," founded in 1792, and dedicated to expanding working class political representation.

one of misery" to ". . . an assertion of the possibility of the harmonious satisfaction here on earth of what are assumed to be normal human appetites." There was a new mass awareness of the scope for economic and political institutions to serve the material needs of all people. Political representation, notably suffrage, was widely seen to be the key. There was a new questioning of established social ranks, famously so in France in the latter part of the eighteenth century. *The Marriage of Figaro* by Beaumarchais (1778) had Parisian audiences siding with the servants in laughing at the aristocracy and deeply questioning their privileges.[19]

The three words that best capture the spirit of the period are "liberté, égalité, fraternité" (liberty, equality, fraternity)—the motto of the French Revolution (and adopted as France's national motto in the late nineteenth century). Although the first few decades after the French Revolution hardly lived up to these lofty words, and the suffrage that emerged was largely confined to men with property, there can be little doubt that the underlying ideas had lasting impact. "Liberty" was understood in a way consistent with modern usage (as in, say, Rawls, 1971), in that the individual was deemed to have whatever freedoms were consistent with like freedoms for others. "Equality" was not, however, understood as equality of outcomes but was defined in terms of legal rights of *opportunity*—that the law must be the same for everyone and so allow all citizens equal opportunity for public positions and jobs, with the assignment determined by ability. There was little immediate sign of a perceived role for the state in redistribution of rewards, although some calls for this did start to emerge in the 1790s with the left-wing Jacobin Club and (in particular) François-Noël (Gracchus) Babeuf.[20] However, if there was hope for poor people in the mainstream ideas of "liberté, égalité, fraternité," then it was more in "fraternity" than "equality"; as Brinton (1934, p. 283) explains:

> *Fraternity had meant to the hopeful eighteenth century the outpouring of its favorite virtue, benevolence, upon all human beings, and especially on the downtrodden and the distant—on peasants, Chinamen and South Sea Islanders.*

[19] For example, in the fifth act, the servant Figaro asks the Count who employs him, "What have you done to deserve such advantages? Put yourself to the trouble of being born—nothing more. For the rest—a very ordinary man!" Although the play was written in 1778, it was censored by King Louis XVI and was not performed until 1784. It is widely seen as a precursor to the French Revolution.

[20] Fleischacker (2004) gives credit for anticipating the modern concept of distributive justice to Gracchus Babeuf, though he also gives credit to the German philosopher Johann Fichte, a follower of Kant. (A seemingly odd pair: Babeuf is considered a founder of Communism and was executed in 1797 for his rebellious left-wing ideas, while the anti-Semitic Fichte is considered a key influence on the National Socialist movement in Germany.) However, de Montesquieu (1748)) appears to have beaten both to the honor.

Similar views were being heard in America where advocates of a strong state role in fighting poverty saw this as an essential element of what it meant to be "a great friendly society" (Alexander Everett, 1827, quoted by Klebaner, 1964, p. 394).

New philosophical and economic thinking from the mid-eighteenth century had opened the way to this Poverty Enlightenment in the last few decades of that century. Significant cracks had started to appear in mainstream views about the role of the state in influencing the distribution of wealth. A key step in this philosophical thinking was the rejection of the view that prevailing inequalities were inevitable. The social contract approach that emerged in the seventeenth century (often attributed to Thomas Hobbes) asked a fundamental question: How should we decide what constitutes good government? In modern terms, this is a question of evaluation, and the relevant counterfactual was a "natural state" in the absence of government. Like all counterfactuals, the natural state was unknown and open to debate.[21] Hobbes argued that it would be a state of conflict, of "all against all." The question was taken up again in the late eighteenth century by Rousseau, who opened up an important new strain of thinking about the distributive role of the state. In his *Discourse on the Origin of Inequality*, Rousseau argued that, although self-interest was a motivation in the natural state, so, too, was empathy for the situation of others.[22] Human institutions, however, can develop to either support or thwart our natural empathy. Rousseau thus saw poverty and inequality as stemming in no small measure (though not solely) from bad institutions—social arrangements that created ". . . different privileges, which some men enjoy to the prejudice of others, such as that of being more rich, more honored, more powerful or even in a position to exact obedience." Here Rousseau made a key step in recognizing the role played by institutions, including governments, in influencing distribution.[23] Poverty was not then inevitable.

Prominent philosophical writings called for respect for poor people as fellow citizens. Kant (1785, p. 62) put forward the idea that every rational human being must be treated "as an end withal, never as means." This was indeed a radical idea, which gave poor people the same moral worth as rich people. Of course there was some measure of respect for poor people even in (say) de Mandeville's earlier writings, but it was a respect for their labor, consistent with the role assigned to them by their birth. They were merely the means to an end. In Kant, by contrast, there was respect for all rational agents, whatever

[21] Rousseau (1754, p. 11) put the point nicely: "The philosophers, who have inquired into the foundations of society, have all felt the necessity of going back to a state of nature; but not one of them has got there."

[22] Rousseau was writing prior to Darwin. Scientific research on animal behavior has revealed strong social and empathic behaviors (de Waal, 2009), suggesting deeper origins for human sociability. It has also been argued that (recently discovered) mirror neurons are the neural foundation of such behavior; see, for example, Keysers (2011).

[23] Rousseau allowed for the existence of what he termed "natural inequality," which would exist in the counterfactual "natural state." Natural inequality reflected innate differences (health, strength, mental ability).

their economic circumstances. This was an essential step for both political equality and comprehensive antipoverty policy, although both were still a long way off.

A longstanding view—often attributed to Cicero in ancient Rome—distinguished justice from beneficence, with only the former entailing a role for the state (Fleischacker, 2004, Chapter 1). Local religious organizations had long been charged with the beneficence role. One crack was opened up by Kant. Theologies have long applauded charity as virtuous. Kant questioned this, arguing that there was an inherently unequal relationship between giver and receiver in charity for poor people; therefore, Kant questioned whether it was "virtuous" to give alms that flatter the giver's pride:

> *Kant sees moral corruption in the private relationships by which well-off people bestow of their bounty to the needy and looks to the state to provide for a more respectful relationship between rich and poor.*
>
> ***Fleischacker (2004, p. 71).***

Such challenges to established thinking about beneficence paved the way for much public debate in Europe and America about the role of the state in fighting poverty and in distribution more broadly, and an eventual shift of responsibilities from religious organizations to the state.

Economic thinking was also advancing. Smith (1776) lambasted the mercantilist view that a country's economic welfare was to be judged by the balance of trade. This had long been questionable (not least for ignoring corrective adjustments through price changes).[24] By arguing for a broader conception of welfare based on the population's command over commodities (including basic consumption goods, not just luxury goods, and also including leisure), Smith opened the way to seeing progress against poverty as a goal for development, rather than a threat to it.[25] Similarly, he argued that higher real wages for workers was a good thing, also in contrast to prevailing mercantilist views (Smith, 1776, Book 1, Chapter 8).

Smith saw the virtue of self-interest—though he did not see it as the sole motive for human behavior (Smith, 1759, Chapter 1, I.I.1)—but only in so far as it advanced social welfare, which depended crucially on the institutional context. And gone was the "utility of poverty," with its negatively sloped individual supply function.[26] Despite the popular characterizations of Smith's noninterventionist views in the twentieth century (Rothschild, 2001), he argued in favor of promotional antipoverty policies, such as limited public subsidies to help cover tuition fees for the basic schooling of the "common people" (Smith, 1776, Book 5, Chapter I, Article 2d). However, on this and other social

[24] See Blaug's (1962, Chapter 2) discussion of Smith and the mercantilist doctrines.

[25] See the discussion in Muller (1993, p. 58). Also see Himmelfarb's (1984a,b) discussion of Smith's views relative to others around the same time.

[26] "Where wages are high, accordingly, we shall always find the workmen more active, diligent, and expeditious, than where they are low" (Smith, 1776, p. 72).

issues, Smith was evidently far more progressive than most of his peers. (Note that Smith was writing at roughly the same time as other thinkers such as Joseph Townsend.)

The changes in popular and scholarly thinking around this time came with implications for ongoing policy debates relevant to income distribution. One such debate was on whether income taxes should be progressive and whose incomes should be taxed.[27] The milieu gave impetus to arguments for redistributive taxation. Smith had strongly favored exempting subsistence wages, as did others subsequently, including those who favored proportional taxes above the exemption—implying a progressive tax system overall.

Another policy debate concerned the distribution of the gains from natural resources, notably agricultural land. In a pamphlet (addressed to government of the new French Republic, but with broader relevance), Paine (1797) argued that agricultural land was "natural property," to which every person had a legitimate claim. There was, nonetheless, an efficiency case for its private ownership. So instead of being nationalized, agrarian land should be subject to taxation—a "ground rent," the revenue from which should be allocated equally to *all* adults in society, as all have a claim to that property. (He also made provision for an additional old-age pension.) And this was (explicitly) not to be seen as charity but as a right. Paine's proposal was a comprehensive antipoverty policy; indeed, it appears to have been the first "basic income scheme"—an idea we return to in Section 22.9, but which has not yet seen national implementation in any country.

An important prelude to the eventual emergence of promotional policies came with new thinking on the importance of *schooling*: "Illiteracy had become a stigma instead of an ordinary accompaniment of humble life" (Brinton, 1934, p. 279). Condorcet, the late eighteenth-century French philosopher and mathematician, advocated free universal basic education (though warning against the state instructing on moral or political matters, as he greatly valued diversity in views); Condorcet also advocated equal rights for women and all races (Jones, 2004). However, these were still radical ideas, well ahead of implementation. The classical economists who came to dominate thinking about policy in the nineteenth century also saw education as having the potential to make economic growth more poverty-reducing, notably by attenuating population growth through "moral improvement." But they did not see mass education as having a role in promoting that growth and saw little scope for mass *public* education (Blaug, 1962, p. 216).

An important contribution of the First Poverty Enlightenment was in establishing the moral case for the idea of public effort toward eliminating poverty. That moral case developed out of a new respect for hard-working poor people, as people, on the part of the elites—what de Waal (2009, p. 116) calls "emotional identification." Important new progressive ideas emerged in the writings of Smith, Rousseau, Kant, Fichte, Condorcet, Babeuf, and others. However, we were still a long way from the three premises identified in the introduction. Although the First Poverty Enlightenment brought

[27] Musgrave (1985) reviews the history of this and other debates in public finance.

about new thinking relevant to antipoverty policies, it did not mark any dramatic change in the lives of the poor, and they were still being blamed for their poverty; the belief that poor people were to blame for their own poverty persisted into the nineteenth and twentieth centuries.[28] Except for relief under the Poor Laws in England and Wales, neither private assistance nor public support for poor people showed any marked rise in Europe, from their relatively low levels (Lindert, 2013). The main economic beneficiaries of the First Poverty Enlightenment were probably in the middle class, who could now aspire to sources of wealth and power they had previously been excluded from.

22.5. THE LONG GERMINATION OF THE IDEA OF A WORLD FREE OF POVERTY

Although mercantilist ideas lost influence with the emergence of classical economics, mainstream thinking in the nineteenth century still held little prospect for a world free of poverty. A new economic growth path had emerged, starting in England toward the end of the eighteenth century, stemming from the technical innovations of the industrial revolution. However, at the time, it was not widely believed among either supporters of capitalism or its critics that workers would share much in this new growth process. (As we will see in Section 22.7, their pessimism on this point was excessive.) Well beyond the start of the industrial revolution, poverty seemed as plentiful as ever. Social novels (such as Dickens's 1838 classic, *Oliver Twist*) and qualitative observational studies (such as Engels', 1845) described the poor health environments and harsh working conditions of English industrial cities in the mid-nineteenth century. Descriptions of working class diets in England around this time suggest levels of living that would almost certainly be considered "poor" in any developing country today (Ravallion, 2015).[29]

The economics of the time appeared to offer little reason to be hopeful about progress against persistent poverty. The classical theories of wage determination allowed the possibility of a short-term rise in real wage rates through an upward shift in the aggregate demand for labor associated with technical progress. However, the induced growth in the size of the working class due to higher earnings—due either to higher fertility or lower child mortality—would soon bring the wage rate back down to the subsistence level. Thomas Robert Malthus is famous for this argument, but a version is also found in Smith (1776, Book 1, Chapter 8). The induced population growth that was central to such Malthusian dynamics was seen to reflect the "moral weaknesses" of poor people. As Sandmo (2013) notes, the idea that population growth would ensure that real wages

[28] See, for example, Klebaner's (1964) descriptions of views of poverty in nineteenth-century America.

[29] For example, Burnett (1969, p. 273) writes that: "The diet of agricultural laborers in mid-century, as of the poorest urban workers throughout the century, consisted essentially of bread—usually white rather than brown because this was more palatable without butter—potatoes, small quantities of tea, cheese and sugar, and meat perhaps once or twice a week."

would stay constant despite technical progress was widely held even to the end of the nineteenth century; see, for example, the writings of Wicksell (1901). The economics was hardly conclusive; lags in the population response and repeated shifts in aggregate demand for labor with technical progress could still yield a secular rise in real wages. The choice-theoretic foundations of the assumed income effect on family size were never clear, but a seemingly common view was that, for poor parents, children were a form of saving for the future. The child wage rate was the return to that saving (net of maintenance costs). A higher wage rate would then be expected to increase the demand for children, thus increasing future labor supply. The classical schema was seen to point to seemingly powerful demographic correctives that would tend to inhibit progress against poverty in a growing economy.

Nor did the most influential classical economists after Smith offer much support for direct public interventions to fight poverty. Indeed, Malthus and David Ricardo were positively hostile to the idea of antipoverty policy, with incentive arguments figuring prominently in their writings. They claimed that such antipoverty policies would discourage work effort and savings and create poverty rather than remove it. Again, the behaviors of poor people were faulted by the elites.[30]

Here, too, it is hardly evident that the economics was decisive one way or the other. Indeed, Malthus (1806) acknowledged that better health and education for working class families could break the brutal population corrective to rising real wages. However, the main interpretation given to the economics of the time was hostile to such policies. In no small measure, this was the intellectual rationalization of a political backlash against the First Poverty Enlightenment, notably among the elites in England who resisted the new liberal ideas that were traveling across the English Channel from France.

22.5.1 The Debate About the Poor Laws

By the early nineteenth century a major public debate about the Poor Laws began (though debates about poverty relief dated back to at least the late seventeenth century). A strong political push for reform came from the landlords, who were financing relief under the Old Poor Laws, who dominated the English parliament around this time, and who were (it seems) no longer worried about impending revolution (Lindert, 2004, Chapter 4). The backlash against the Poor Laws often invoked incentive arguments, and England's classical economists were widely cited by critics of the Poor Laws, including those from America (Klebaner, 1964).

This was a significant debate in the history of thought on poverty. For some time, powerful critics had been concerned about the overall cost of the policy. Labor migration in response to industrialization had meant that local landlords were left to finance a rising

[30] Klebaner (1964) points to official claims that 75–90% of pauperism in the United States in the nineteenth century was due to intemperance. Also see Burnett (1969, pp. 274–276) on similar arguments in England.

support bill for children and the elderly (Solar, 1995). Nor was work found by all, and unemployment was causing many in Europe and America to turn to the state for help. But these were not the explanations for the rising relief bill that gained favor. Observers such as de Tocqueville (1835) (in a memoir reporting on a visit to England with the aiming of understanding why there were so many paupers despite the country's affluence) argued that the Poor Laws were a disincentive to work, such that they helped create the poverty problem they aimed to solve. Prominent classical economists, including Malthus (1806, Chapters 5 and 6) and Ricardo (1817), argued for either abandoning the Poor Laws or at least reforming them to ensure better targeting.[31] In an influential earlier pamphlet, *A Dissertation on the Poor Laws*, Townsend (1786, p. 17) wrote that "[t]hese laws, so beautiful in theory, promote the evils they mean to remedy, and aggravate the distress they were intended to relieve." Assumptions about incentives were the core of Townsend's argument. Public relief from chronic hunger would discourage work, and the fiscal burden on the landholding class would discourage the growth of manufacturing and innovation in agriculture (Townsend, 1786, Section V). Ricardo (1817, p. 61) predicted (plainly with huge exaggeration) that the cost of the Poor Laws would rise out of control, that "whilst the present laws are in force, it is quite in the natural order of things that the fund for the maintenance of the poor should progressively increase until it has absorbed all the net revenue of the country." Malthus argued that the Poor Laws encouraged early marriage and high fertility (though counterarguments could also have been made that ensured old-age support would reduce fertility). Moral hazard appears to have been a concern, whereby assistance to those who took high risks, and lost out, would encourage excessively risky behavior. The Poor Laws came to be seen by many as a cause of poverty rather than its cure. Similar debates were also being waged about America's poor laws, with calls for reforms to cut their rising cost (Klebaner, 1964).

However, the evidence was clearly weak for the claims that behavioral responses to the laws were an important cause of the poverty they tried to address. The evidence appears to have been largely based on easily manipulated anecdotes and characterizations, with plainly weak claims of attribution; for example, was the claimed high incidence of intemperance a cause or effect of poverty? Nor was there much recognition that nonintervention could be socially costly, too—that problems of heterogeneous risk and asymmetric information could entail that the private insurance was unavailable,[32] and that uninsured risk could spill over into production and investment decisions of poor people in ways that could impede long-term prospects of escaping poverty. For example, against the concerns that relief would reduce the labor supply, Solar (1995) argues that the Old Poor Laws had the opposite effect, by providing security against the risk of

[31] See the discussion of the views of Malthus and Ricardo on this topic in Sandmo (2013).

[32] This economic argument for social insurance was not well developed in the literature until much later, notably by Rothschild and Stiglitz (1976).

unemployment for smallholders who were considering whether to become laborers instead. The type of model, outlined in Section 22.2 (Figure 22.1), motivates social protection even for people at their "high" steady-state equilibrium (point C in Figure 22.1). For example, imagine someone at that equilibrium receiving a sufficiently large negative shock to push them just past the unstable equilibrium. There will be no chance of recovery, and destitution will be the inevitable result. Lack of insurance could well have been a more important reason for poverty than too much insurance.

Although incentive effects and dependency were a legitimate concern, the economic arguments against England's Old Poor Laws may well have been exaggerated to serve political ends (and it was not the first or last time this happened). The "evidence" was weak, and the arguments were somewhat one-sided, with many potential economic benefits of the laws ignored.

Significant reforms to the Poor Laws were implemented in 1834 (including repeal of Speenhamland). Spending was slashed, from a peak of about 2.5% of national income around 1830 to 1% in 1840 (Lindert, 2013, Figure 1). Wider use was made of workhouses. These had long existed, and by the late eighteenth century, 1–2% of the population of London was seeking relief in some 80 workhouses.[33] Their role expanded under the reform effort to ensure better targeting, and the new nineteenth-century workhouses appear to have been even more unpleasant and punitive places than in the past (described well in *London Lives*). Earnings in the workhouse were never to exceed local wages (Beaudoin, 2007, p. 80). The policy became better targeted, but it lost the broad public support of the Old Poor Laws and (indeed) became the subject of intense social criticism. By confining beneficiaries to workhouses, the reformed policy was seen by critics to treat poor people as criminals. The conditions under which inmates were kept became a specific focus of criticism, famously so in the early chapters of Dickens's (1838) *Oliver Twist*. And the criticisms (which started almost immediately) of the New Poor Laws were not just confined to social critics but reached deeply into the leading circles of the Conservative Party, including Benjamin Disraeli (Himmelfarb, 1984a,b).

22.5.2 Utilitarianism

Social contract theory, with its emphasis on rights and freedoms, lost ground in the nineteenth century to a rival school of thought, utilitarianism. This also emerged in the late eighteenth century and over the next 200 years came to have great influence on normative economics—indeed, it became the "official theory of traditional welfare economics" (Sen, 2000, p. 63). Jeremy Bentham, the founding father of utilitarianism, was motivated

[33] See the entry on workhouses in *London Lives 1690–1800*. Also see in Hindle's (2004, p. 176) discussion of the use of encouragements to work under the Old Poor Laws, whereby the church vestry often became a "job-creation service" (p. 176). Workhouses existed elsewhere in Europe, including Holland where they were introduced in Amsterdam around 1600 (Beaudoin, 2007, p. 48).

by practical policy reform, and this led him to reject ideas like "natural rights." (Artz, 1934, p. 83, quotes him as describing the *Declaration of the Rights of Man and of the Citizen* as "a hodge-podge of confusion and absurdity.") Instead, utilitarianism advocated that social choices should maximize the sum of utilities across all individuals, where "utility" was equated with "happiness." Assuming diminishing marginal utility of income, this objective generated a case against *income* inequality because the marginal losses to rich donors of any mean-preserving transfer would be outweighed by marginal gains to poor recipients. This did not, however, open the floodgates of redistributive interventions. Assuming diminishing marginal utility of income and a common utility function only implied that equality of incomes was optimal if total income was invariant to its distribution. The case was unclear if income redistribution lowered overall output, as Bentham expected to be the case. Even aside from incentive effects, merely introducing interpersonal heterogeneity (such that the utility valuation of a given income level varies) upsets the claim that an equal allocation of income maximizes social welfare though this point did not seem to get the same attention as the growth-equity tradeoff.

Bentham and followers had seen government as a necessary evil, and put any actual or contemplated policy effort to the utilitarian test. Some of the literature has (derisively) characterized this as a period of "laissez-faire," although to an economist's eyes it was a welcome discipline in sound policy making, to ensure maximum social welfare. The real issue was what one meant by "social welfare." The influential rights-based thinkers on policy prior to the utilitarians, such as Condorcet, would no doubt have also advocated higher social welfare but would have rejected any attempt to equate welfare with "happiness" or "utility" (Rothschild, 2001).

By the mid-nineteenth century, it was becoming accepted in prominent progressive circles that the state did have a role in "... redressing the inequalities and wrongs of nature" (Mill, 1848, p. 805). Even so, it is clear that poverty was still widely accepted as a normal state of affairs. Poor people were still being blamed for their poverty (notably by their reproduction), and there was little role for the state. Even protection was increasingly "targeted" to extreme cases. The best that could be hoped for was that workers would somehow come to see the wisdom of curtailing their desired family sizes. Even among the most progressive utilitarian voices of the time (such as John Stuart Mill), the closest one came to promotional policies would be to point to a role for education of the working class in reducing population growth, but with a strictly limited role for the state.

22.5.3 Schooling Debates

Children from poor families typically started their working lives at an early age; although the evidence is patchy, it was common prior to the mid-nineteenth century for working class children in England to start looking for work from 7 years of age (Cunningham, 1990). The survival of the family often demanded that every able-bodied person worked.

Any skills required would only be those that could be passed on by the family. Idle poor children were abhorrent to the rich; work was the only solution. Child labor was not only condoned but widely seen as desirable; unemployment of poor children was believed to be the bigger social problem (Cunningham, 1990). The idea of mass public schooling appears to have had little support. Indeed, echoing de Mandeville's views, a common view was that mass schooling was wasteful and even dangerous. By the middle of the nineteenth century some 40% of children aged 5–9 in England and Wales were still not in school.[34]

Nor was the state deemed to have an important role in the schooling that was provided. Before the nineteenth century, and well into that century in some countries (including England), almost all schooling received by children from poor families was provided by religious groups. The system of voluntary schooling in England and elsewhere in Europe was clearly highly stratified and unequal. Schooling by religious groups had a mixed record. In England, the church resisted any public role in provision yet also left much unmet demand (Lindert, 2004, Chapter 5). The debate on mass schooling opportunities continued in England until quite late in the nineteenth century, and the country lagged behind much of Europe and North America in schooling attainment, despite its wealth.

Poor families did not always see Church schools as being in their interests. Informal private schools were often more promising for those who could afford them. Van Horn Melton (1988, p. 11) describes the "backstreet schools" in Austria and Prussia that offered more efficient instruction "subordinating religious instruction to the goal of imparting literacy to their pupils," and it appears that these schools were often favored by poor parents who were eager to ensure their children's efficient learning and eventual employability; with reference to Prussia, Van Horn Melton (p. 11) writes that: ". . . backstreet schools offered poorer families a more cost-effective means of acquiring literacy." This echoes observations of the "backstreet schools" found throughout India today, reflecting evident failures of the state-run schooling system (Probe Team, 1999).

A change in popular views about schooling for poor families started to be evident in much of Europe and North America in the mid-nineteenth century. Mercantilism had lost its influence, and the classical economics that replaced it was not opposed to promotional policies such as public schooling—policies that were capable of a propoor change in the distribution of wealth. The working conditions of children in the factories of the time provided fuel for social novels *and* for the increasingly vocal critics of capitalism, most notably Karl Marx and Friedrich Engels. Prominent calls started to be heard for improving the working conditions of children and for schooling as the better way to

[34] This is based on the 1851 census (as reported in Cunningham, 1990, Table 1); 39% of boys and 44% of girls in this age group were not classified as "scholars" (the alternatives being employment or "at home"). The subsequent spread of literacy was also highly uneven geographically (Stephens, 1998).

address unemployment. Schooling for poor children came to be seen as key to their self-improvement and mobility. Mass schooling was also believed to have external benefits, such as reduced crime rates.

National legislation for compulsory schooling had only emerged in a few countries (including Austria and Prussia) toward the end of the eighteenth century but was becoming widespread in Europe and North America by the late nineteenth century.[35] This followed a protracted public debate in Britain, Europe, and North America during the nineteenth century (Weiner, 1991, Chapter 6). Although there were some who argued against almost any intrusion by the state into private decision making,[36] this does not appear to have been the main argument of opponents. Mill's (1859) influential volume *On Liberty* argued that the state had a role in compelling parents to school their children, although Mill did not favor government monopoly in the provision of that schooling. Opponents had long argued that schooling the poor would lead them to unrealistic aspirations (Vinovskis, 1992). As one would expect, the industries that were heavily dependent on child labor lobbied against compulsory schooling although over the course of the nineteenth-century industrial capitalists became more supportive of mass schooling because they wanted to create the more skilled workforce needed for the new technologies (Bowles and Gintis, 1976, with reference to the United States). However, this was not simply a matter of schooling catering to the needs of new technologies developed under capitalism; the debates about schooling were broader socially, and it is not clear that industrial capitalists had that much influence (Vinovskis, 1992). Poor parents and local communities were also increasingly vocal in their demands for mass public schooling. It seems that by the latter half of the nineteenth century the earlier unrealistic aspirations of poor parents for a better life for their children had started to become more realistic. There were also administrative constraints on enforcement to overcome; it was not until birth registration systems had been developed around the mid-nineteenth century that truancy laws could be properly enforced (Weiner, 1991, p. 121).

22.5.4 Socialism

Landauer (1959) identifies the widespread acceptance of poverty in the nineteenth century as one of the factors that led to the emergence of socialism. The leading school of socialist thought, Marxism, saw the root cause of poverty, and most other ills, to be capitalism itself. There was little scope for effective antipoverty policies *within* a capitalist economy; only communism could reliably eliminate poverty. Nor was much value

[35] There were some progressive local initiatives for mass schooling, such as Massachusetts in the late seventeenth century (Weiner, 1991, Chapter 6).

[36] In the United States, one occasionally hears arguments that compulsory schooling is unconstitutional, the reference being to the antislavery amendment introduced near the end of the Civil War on the grounds that (it is claimed) compulsory schooling is "involuntary servitude." See here, for an example.

attached to past philosophical and economic thought on poverty. For example, Marx was as disparaging as Bentham about talk of "rights."[37]

Even so, it is notable that at least a couple of the demands outlined in the *Communist Manifesto* of Marx and Engels (1848) can be recognized today as quite mainstream antipoverty policies, including progressive income taxation and free education in public schools. Fleischacker (2004) identifies one key influence of Marx's thinking on subsequent non-Marxist thinkers, including Rawls, namely, his insistence that human nature was largely a product of social context. Instead of seeing poverty as the outcome of individual attributes (being lazy is a favorite in some quarters), one should look to social influences on behavior. Of course, this idea also had pre-Marxian antecedents, notably in Rousseau.

22.5.5 Social Research

Much new research on social problems was emerging in the nineteenth century, and poverty was increasingly seen as a social problem. Social research was used to promote better informed public debate on antipoverty policy. Important contributions included Eden's (1797) three-volume tome on poverty in England and Wales in the late eighteenth century, Mayhew's (2008) newspaper reports on London's poor in the 1840s, Frederic Le Play's budget studies of working class families in Europe in the mid-nineteenth century (Brooke, 1998), Mathew Carey's use of data on budgets and wages of poor people to "startle the complacent into giving alms" in Philadelphia in the 1830s (Klebaner, 1964, p. 384), and the work of the German statistician Engel (1857), who studied the relationship between household food expenditures and income, in which he found what came to be known as Engel's Law, namely, that the poorer a family is the higher the share of its budget devoted to food.

Landmarks in the development of modern scientific research on poverty were the (largely independent) studies by Booth (1903) and Rowntree (1902), which documented the living conditions of England's poor (in London and York, respectively) in the late nineteenth century. These were pioneering measurements using seemingly careful household surveys that revealed to nonpoor people how poor people lived. Their work attracted much attention.[38] The English public was shocked that one million Londoners—about one-third of the population—lived below Booth's frugal poverty line of 21 shillings per week for a family. This news came after a period of rising real wages, which added to the shock. Nor could it be said that this was too generous a poverty line.

[37] Fleischacker (2004, p. 97) quotes Marx as calling appeals to rights "ideological nonsense."

[38] Booth is often credited with inventing the poverty line. There were also antecedents to the idea of the poverty line in Booth and Rowntree, including the "standard of comfort" proposed by Davies 100 years earlier (Allen, 2013).

By my calculations it was equivalent to 1.5 lb of good wheat per person per day—a frugal line, not very different from (say) India's poverty line in the 1990s.[39]

Booth's research responded to a demand for clarity and data among legislators. His empirical research into old-age poverty and its geographic variation influenced Britain's introduction of a public pension in 1908 (Thane, 2000, Chapter 9) and national insurance in 1911 (Himmelfarb, 1984a,b). The research of Booth and Rowntree also stimulated debates about poverty. For example, 15 years after Booth's books appeared, Alfred Marshall argued that there was even more poverty in Germany than Booth's figures suggested was the case in England; this was in response to Marshall's (1907, p. 12) perception that "one of the few things which every German knows for certain about England is that there are a million people in London living in extreme poverty on the verge of hunger."

The close observational studies of poverty by Booth and Rowntree were influential in social science research. Hunter (1904) followed their lead in studying poverty in the United States. Village studies in India by Mann and collaborators were also influenced by Booth and Rowntree (Thorner, 1967). A long and distinguished tradition of quantitative-economic studies of selected villages followed, including surveys by Askok Rudra and Pranab Bardhan (Bardhan, 1984a), Bliss and Stern (1982), Walker and Ryan (1990), and Lanjouw and Stern (1998). Booth's approach influenced the development of quantitative sociology in both Britain and the United States.[40] Townsend's (1979) empirical study of poverty in England some 80 years later clearly owed much to Booth and Rowntree. So, too, did the Chicago School of Sociology that began studying urban poverty in the United States during the 1930s.

The late nineteenth century saw new questioning of the longstanding idea that poverty was inevitable in any capitalist economy and the emergence of prominent arguments for promotional antipoverty policies in such an economy. Although the late eighteenth century gave birth to the modern idea of distributive justice, it was not until the late nineteenth century that we saw the emergence of the idea of a world free of poverty. By then it had become widely accepted among the "cultivated circles" that a trend rise in the real

[39] Marshall (1907) estimates that 21 shillings was equivalent to three-quarters of a bushel of good wheat. At 13.5% moisture by weight, a bushel of wheat weights 60 lb according to the Wikipedia entry on "bushel." I assume a household of 4.5 people, which is the lower bound of the range 4.5–5 given by Booth (1903, Chapter 4) for the average size of working men's families at the time. Booth's line is thus equivalent to slightly less than 700 g of wheat per person per day. Of course, this is just the wheat equivalent. A reasonable dietary breakdown would be 400 g per person for wheat and the remainder for meat, vegetables, and (very minimal) nonfood needs. This then is similar to India's national poverty line in 1993, which World Bank (1997) calculates to be equivalent to a daily food bundle per person of 400 g of coarse rice and wheat, 200 g of vegetables, pulses, and fruit, plus modest amounts of milk, eggs, edible oil, spices, and tea. After buying such a food bundle, one would be left with about $0.30 per day (at 1993 purchasing power parity) for nonfood items.

[40] On Booth's influence see the Wikipedia entry on Charles Booth and the Archive maintained by the London School of Economics.

wage rate was a sign of overall progress (Daniels, 1898, p. 203). The historian Webb (1974, p. 384) argues that in late nineteenth-century England it came to be recognized that poverty "could and must be eliminated."[41] Near the turn of the century, Marshall's (1890, p. 2) *Principles of Economics* was posing the question quoted at the beginning of this chapter, bemoaning that the children of the poor received too little schooling (p. 467), and sketching policies for fighting poverty (especially., pp. 594–599) that were not just intended as short-term moralistic palliatives but were driven by a recognition that persistent poverty was itself a constraint on wealth generation. Marshall (1890, p. 468) wrote of the "cumulative evil":

> *The worse fed are the children of one generation, the less they will earn when they grow up and the less will be their power of providing adequately for the material wants of their children; and so onto following generation.*

Thus:

> *The inequalities of wealth, and especially the very low earnings of the poorest classes . . . (are) . . . dwarfing activities as well as curtailing the satisfaction of wants (p. 599).*

Marshall's reference here to "dwarfing activities" anticipates a view that is prominent in development thought today whereby certain inequalities are seen as instrumentally important inhibitors of overall economic progress, notwithstanding their intrinsic relevance in "curtailing the satisfaction of wants." Although Marshall was careful to avoid naïve utopianism (see, especially, the comments in Marshall, 1907), his writings reflect a far more optimistic perspective on social policy as a means of expanding opportunities for all to share in the potential of a competitive market economy. Here we had a forthright and prominent advocacy of promotional policies such that ". . . children once born into it [poverty] should be helped to rise out of it" (p. 598).

Importantly, this new optimism was starting to be shared by poor parents, who raised their demand for schooling for their children. By the late nineteenth century, it seems that most poor parents in Europe and North America were anticipating that their children would encounter better economic opportunities than they had. Helped by significant medical and public health advances that were improving child survival rates and raising life expectancy, investing in children's schooling was seen as a far less risky than it had been early in that century (and before then) when the children of the poor working class had little real hope of being anything else than working class and not much chance of being less poor workers than their parents. The demand for mass schooling thus rose along with the supply. Parents were still investing in their children to help secure their own future welfare (formal social security systems were not yet available), but they were investing more in the quality of those children. Fertility rates were falling.

[41] Beaudoin (2007, p. 100) gives the idea a more recent origin in the twentieth century, after World War II.

After the First World War, there was a mounting enthusiasm for policy intervention in the West, and there appears to have been broad agreement that greatly reducing, if not eliminating, poverty was a legitimate role for government (Mencher, 1967). In the writings of prominent economists, such as Pigou (1920, Part IV, Chapter 1), it had become accepted that losses to the "national dividend" could be justified by gains to poor people. The incidence of absolute poverty had come to be recognized as an important yardstick for measuring social progress. For example, the eminent statistician Bowley (1915, p. 213) wrote that:

> *There is perhaps, no better test of the progress of a nation than that which shows what proportion are in poverty and for watching the progress the exact standard selected as critical is not of great importance, if it is kept rigidly unchanged from time to time.*

From around the turn of the twentieth century, statistics was being applied to various social issues, including measuring poverty and inequality. A key methodological issue was whether one could rely on sample surveys (instead of doing a census) and how the sampling was to be done (the choice being between purposive and random sampling). Statisticians such as Bowley, Ronald Fisher, and Jerzy Neyman advanced the theory of statistical inference based on random sampling, although it took a few decades before this became common practice for social and economic surveys.[42] Poverty measurement was a leading application and, in due course, sampling methods were to revolutionize the collection of systematic survey data on poverty and inequality by national statistics offices across the world.

By the interwar period it seems that poverty was no longer being seen in mainstream circles as primarily caused by the bad behavior of poor people, but as reflecting deeper economic and social problems. If nothing else, the observation of mass involuntary unemployment during the Great Depression made that clear. And the observations were carried with force to a broad audience through various media.[43] The period saw massive relief efforts (such as the New Deal in the United States). But these were largely transient efforts for protection rather than promotion (Heclo, 1986).

22.6. THE SECOND POVERTY ENLIGHTENMENT

The period from about 1950 saw a new trajectory of more rapidly declining incidence of absolute poverty in the world, as judged by the standards of what poverty means in the poorest countries.[44] From about the same time, a significant shift in thinking was

[42] The two-stage sampling method introduced by Hansen and Hurvitz (1943) was to prove especially useful for countries at all stages of development. On the history of survey sampling methods up to the present, see Bethlehem (2009).

[43] The photos and text of Agee and Evans (1941) describing the living conditions of Southern tenant farmers in the United States in the mid-1930s was an example.

[44] This is shown in Ravallion (2015), drawing on the estimates made by Bourguignon and Morrisson (2002) and Chen and Ravallion (2013).

underway, with bearing on antipoverty policy. This was the Second Poverty Enlightenment, dating from about 1960. Across the globe—including in the newly free countries of the developing world—there was new optimism among policy makers about the scope for fighting poverty. Evidence for the change in public attention to poverty can be found in the striking rise in the incidence of use of the word "poverty" in the writings of the time after 1960. This is evident if one enters the word "poverty" in the *Google Books Ngram Viewer* (the *Viewer* hereafter) (Michel et al., 2010). The *Viewer*'s counts are normalized by the total number of words that year, giving the "incidence" of that word. The upturn in incidence started around 1960. By 2000, the incidence of references to "poverty" reached its highest value in 300 years. And the rise in incidence continued after 2000 and up to the latest year (2008) for which the data are available at the time of this writing; indeed, with moderate smoothing of the time series, in 2008 poverty had the most attention in the literature since 1600.[45] Attention to poverty appears to be higher now than any time since 1800 while the incidence of extreme absolute poverty is at its lowest point since then (Ravallion, 2015).

Similarly to the First Poverty Enlightenment, the Second was a time of radical questioning and instability, although, unlike the First, it did not come in the wake of rising absolute poverty. There were demands for new freedoms across the world. There was social ferment and civil unrest in the rich countries of the world, and newfound political independence combined with much political and economic upheaval in the poor countries of the world.[46] Also similarly to the First Poverty Enlightenment, there was new scholarly thought that had great bearing on antipoverty policy.

In philosophy and economics, the 1960s and 1970s saw renewed questioning of the utilitarian paradigm as a basis for public action against poverty and inequality, and in other domains of public policy. Critics of utilitarianism questioned whether policies that entailed welfare losses to the poorest could ever be justified by sufficiently large gains to the richest. A case was made for the ethical prioritization of helping the poorest first, as in Rawls's (1967, 1971) formulation of the principles of justice, which we return to below. The 1970s saw efforts to generalize the utilitarian schema by embodying an

[45] The relevant plot up to 2008 and as far back as possible can be found here. There are two spikes in 1634 and 1659. Naturally, the volume of words in the *Viewer*'s database is low in these earlier years, often with only a few books per year. Each of these spikes largely reflects one or two volumes that used the word "poverty" a lot. This is clearly deceptive. With any smoothing parameter greater than three, the peak year becomes the last year in the series, 2008. Also note that the count is case sensitive. The use of capitalized words mid-sentence was more common in English writing of the seventeenth and eighteenth centuries so it is important to include capitalized words when going back that far. But this matters little after 1800 or so.

[46] Although the 1960s was a famous period in the West, with vocal new movements for peace and racial and gender equity, much was also happening in the developing world. In the 1960s alone, 32 countries in Africa gained independence, though often with contested borders. China's "Cultural Revolution" started in 1966 and wreaked havoc for 10 years. South Asia (Bangladesh and India) and parts of Africa were fighting famines in the 1960s and 1970s, and there was much political instability; even relatively stable India had its share of political upheaval including the "Emergency" in the mid-1970s.

aversion to inequality of utilities, such that the marginal social welfare attached to higher utility fell with the level of utility. In principle, marginal social welfare could then be driven down to virtually zero at a sufficiently high level of utility. Once one made the extra step of allowing the possibility that marginal social welfare could go to zero above some point, prioritizing poverty reduction could be interpreted as the negative of social welfare maximization.[47] Whether or not one took that extra step, there was clearly common ground in these different emerging schools of thought about the social welfare objectives of public policy.

For many economists, the more contentious step (and it is still contentious) was attaching intrinsic value to "rights" and "freedoms." Dissatisfaction with the lack of attention in economics to popular concerns about individual rights and freedoms was evident during the Second Poverty Enlightenment. Of course, the freedom to trade freely was often given high value in economics, but this was an instrumental value—the virtue of competitive exchange was a derived one from longstanding Benthamite or Paretean formulations of policy objectives. The scope for ethically contestable policies was evident if one did not put certain rights above all else.[48] Motivated by such concerns, mainstream thinking about poverty in both scholarly and policy circles was being given to nonutilitarian formulations that put freedom as the central issue, most notably in the writings of Sen (1980, 1985, 1999). The idea that poverty is fundamentally a lack of individual freedom to live the life one wants—a severe deprivation of basic capabilities in Sen's terms—and that such freedom has an overriding ethical merit can be traced back to the Second Poverty Enlightenment.

Many policy issues, including debates on antipoverty policies, call for some form of interpersonal comparison of utility. Yet, in the wake of an influential book by Robbins (1935), the period up to around 1950 saw economists striving to purge welfare economics of interpersonal comparisons—leaving little scope for normative economic analysis of poverty or income distribution more generally.[49] One turning point in thinking on this issue came with Arrow's (1951) famous theorem.[50] In due course, Arrow's theorem and

[47] This interpretation is discussed further in Ravallion (1994a), which shows that on introducing inequality aversion into the measure of poverty and allowing for measurement errors in the data on individual economic welfare, the resulting formulation of the objectives of policy in terms of minimizing poverty can essentially be made as close as one likes to the negative of a generalized utilitarian social welfare function.

[48] An example is the various coercive efforts made to encourage poor parents in developing countries to have fewer children; see the examples described by Hartmann (1987).

[49] For an authoritative overview of this and other issues in the history of thought on income distribution, see Sandmo (2013).

[50] Developing arguments first made by Condorcet in 1785, Arrow (1951) established that, under seemingly defensible axioms, a unique social ordering over three or more options that is derived solely from a set of unrestricted individual orderings must be imposed externally.

the work on social choice theory that it stimulated led to a reaffirmation of the need for some form of interpersonal comparability in discussing issues such as antipoverty policy.[51] Ethical considerations soon returned in full force to policy analysis by economists, although it also came to be understood that not all such analyses required fully comparable cardinal utilities (Sen, 1970b). The futility of attempting to infer uniquely comparable utilities solely on the basis of demand behavior also came to be accepted (especially following Pollak and Wales, 1979). The 1970s and 1980s saw new efforts to put poverty and inequality measurement on firmer theoretical foundations.[52] There was an explosion of interest in the measurement of poverty and inequality, both in theory and in practice, starting from around 1970 (Ravallion, 2011).

Other seemingly sacred elements of economics started to be questioned, including whether people were rational, although some of the claims of "irrationality" that emerged from behavioral economics appeared to stem more from limited characterizations of utility functions and/or limited allowances for mistakes (Saint-Paul, 2011). Even the idea that social welfare had to be strictly increasing in all utilities (the Pareto principle) was being questioned as either a sufficient or a morally compelling basis for policy making (as in, for example, Nath, 1969). The Pareto principle was even found to be inconsistent with seemingly mild requirements for personal liberty (Sen, 1970a).

The 1970s also saw a deeper questioning of the efficiency of competitive market allocations. The term "market failure" (introduced by Bator, 1958) had become widely used, and labor and credit markets' imperfections in particular came to be seen as key to understanding poverty. The idea that labor markets were competitive, such that wage rates adjusted to remove any unemployment, had been in doubt since the Great Depression. In understanding poverty in rich countries in the 1960s, the idea of dual labor markets became prominent, following in particular Doeringer and Piore (1971). One segment of the labor market has high wages and good benefits while the second has low wages and little in the way of benefits. Bulow and Summers (1986) showed how this could be an equilibrium given the existence of high costs of monitoring work effort in certain activities, which become the high-wage segment in which profit-maximizing firms pay wages above market-clearing levels (following Shapiro and Stiglitz, 1984). Other activities with low monitoring costs form the competitive segment, which is where the working poor are found.

In another strain of the literature of this period, Akerlof (1970) showed how credit (and other) market failures can arise from asymmetric information, such as when lenders are less well-informed about a project than borrowers, thus constraining the flow of

[51] See the discussion in Roemer (1996). Notice, however, that allowing interpersonal comparisons is only one of the possible resolutions of Arrow's dictatorship result (Sen, 1970a,b).

[52] Important contributions came from Watts (1968), Atkinson (1970, 1987), Kolm (1976), Sen (1973, 1976), and Foster et al. (1984).

credit. This helped explain the efficiency role of institutions and governments in facilitating better information signals and broader contract choices. For example, the idea of asymmetric information gave a new perspective on why share-cropping existed (Stiglitz, 1974). Since the work effort of tenants is unobservable by landowners, an optimal contract strikes a balance between risk sharing and incentives for work. Thus, risk is shared between the two parties.

The new economics of information held important implications for understanding poverty. In a perfect credit market, even poor parents will be able to borrow for schooling—to be paid back from children's later earnings. However, if poor parents are more credit constrained than others, then we will see an economic gradient in schooling, whereby the children of poor parents are less schooled.[53] This is indeed what we see, almost everywhere. There will be too much child labor and too little schooling in poor families. Thus, poverty will persist across generations. Risk market failures can have similar implications. Parents will under-invest in their kid's schooling when they cannot insure against the risk of a low economic return from that schooling.

In due course, this new strain of economic thinking would point to important ways in which inequalities in the initial distribution of wealth could persist and impede overall economic progress; Section 22.8 returns to this issue. The economics also pointed to the scope for promotional antipoverty policies—policies that essentially aimed to compensate for the credit and risk market failures, such as by compulsory schooling laws and public support for schooling, especially for children from poor families. Section 22.9 will return to such policies.

22.6.1 Rawls's Principles of Justice

If there is a single philosophical landmark of the Second Poverty Enlightenment, it must be Rawls's (1971) *Theory of Justice*. Borrowing from early formulations of social contract theory (back to Hobbes), Rawls proposed that the principles of justice should be the social contract agreed to among equals in a veil of ignorance about where they would find themselves in the real world. (The veil of ignorance was a thought device to ensure that morally irrelevant—inherited or acquired—advantages in the real world did not color judgments about distributive justice.) Rawls argued that two principles would emerge. First, each person should have an equal right to the most extensive set of liberties compatible with the same rights for all; this borrowed the idea of liberty that had emerged in the late eighteenth century, famously so in the French Revolution. Second, subject to the constraint of liberty, social choices should only permit inequality if it was efficient to do so—that a difference is only allowed if both parties are better off as a result; this is what Rawls called the "difference principle."

[53] This was postulated in an important economic model of how poverty could persist; see Loury (1981).

This second idea was more radical in its egalitarianism than the French Revolution's motto. However, it was not the kind of radical egalitarianism that said that equality always trumped efficiency. Indeed, society A, with a great deal of inequality, would be preferred by this moral principle to society B, with no inequality, if the poorest were better off in society A. Thus, the principle amounts to maximizing the advantages of the worst off group and hence became known as "maximin." This was explicitly not a proposal to maximize the lowest income, as it is sometimes interpreted, but rather to maximize the welfare of the worst off group in society. The "worst off" people were to be identified by what Rawls called their command over "primary goods." These are all those things needed to ensure that one is free to live the life one wants. This is a broader category than what are often called "basic needs" as it includes social inclusion needs and basic liberties—in short, rights as well as resources.

As Rawls recognized, one will need an index for determining the least advantaged. Possibly because of his evident desire to break all ties with utilitarianism, Rawls avoided using the term "utility function" (or "welfare function"), but this is evidently what he has in mind in his discussions of the "index problem" (especially, Rawls, 1971, pp. 90–95)—namely, a function that expresses the accepted tradeoffs. And Rawls agreed that it is also compelling that those tradeoffs be consistent with individual preferences over primary goods (Rawls, 1971, p. 94). However, he argues that we need not be concerned with the preferences of the nonpoor under the assumption that their primary good vectors are bound to dominate those of the poor.[54] (This is an empirical question, but a plausible assumption in the absence of data.) Thus, the utility function of the worst off person should be decisive in aggregations across primary goods.[55]

The Second Poverty Enlightenment had intellectual roots in the First. Rawls saw his difference principle as an interpretation of "fraternity" (as in the French Revolution's motto): "[T]he idea of not wanting to have greater advantages unless this is to the benefit of others who are less well off." This was a natural step (though it took a long time) from the aspirations for fraternity in the First Poverty Enlightenment. Utilitarianism was seen to be in conflict with fraternity as it could justify losses to the individual in the name of total utility. There would always be some gain to the richest person that could justify a loss to the poorest. The individual is subordinated to the common good, as measured by the sum of utilities. This Rawls rejected.

Rawls saw his theory as a reinterpretation of Kant. Poor people should have the right to veto any scheme that brings gains to the well-off at their expense. In direct

[54] More generally, the partial ordering of vectors of primary goods required by Rawls's maximin principle need not require a mathematically precise aggregation function; a sufficient partial ordering may be possible by only specifying certain generic properties of that function; for further details, see Atkinson and Bourguignon (1982).

[55] "The only index problem that concerns us is that for the least advantaged group" (Rawls, 1971, p. 93).

contradiction to the dominant view 200 years earlier, poverty for some was judged to be unacceptable as the means to others' prosperity. Utilitarianism (by contrast) could not guarantee a satisfactory minimum.[56] And only if a satisfactory minimum was ensured would the social contract be "stable" in that "the institutions that satisfy it will generate their own support" (Rawls, 1967).

The reasoning here was that, as long as the worst off group was happy with the social arrangement, then the rest (all doing better than the worst off group) would have nothing to complain about (Cohen, 1989b). Of course, this reasoning is questionable in the real world as those not in the poorest stratum could be expected to have a different counterfactual in mind when assessing any policy to that of being the worst off. But recall that the social contract was being formed in the absence of information about real-world positions. Rawls argued that maximin was more likely to emerge from rational choice behind the veil of ignorance.

Rawls's theory of justice has stimulated much debate. Harsanyi (1975) questioned whether maximin was a more plausible choice for a social contract than maximizing average utility even behind the veil of ignorance. Roemer (1996, Chapter 5) also questioned whether maximin would emerge as the solution. These critiques rested on the assumption that agents behind the veil would maximize expected utility, which depends solely on their own consumption (and leisure). This requires that subjective probabilities can be assigned to all states behind the veil, which Rawls (1971) questioned.[57] Introducing social preferences could also upset these critiques.

Other critiques of Rawls's theory emerged. Soon after the publication of *Theory of Justice*, Nozick (1974) published a libertarian critique. Nozick gave primacy to historical property rights above all else, although it was never clear on ethical grounds why property rights were never to be questioned.[58]

Sen (1980) took issue with Rawls's concept of primary goods, arguing that this idea does not adequately reflect the freedoms that people have to pursue their goals, recognizing the heterogeneity in the ability of people to transform primary goods into freedoms. This critique led to Sen's (1985) conceptualization of welfare in terms of primary "capabilities"—"what people are able to be and do (rather than in terms of the means they possess)" (Sen, 2000, p. 74).

As Pogge (1989) argues, one can defend the key aspects of Rawls's principles of justice without accepting his rationale in terms of a social contract. Roemer (2013) argues for a

[56] Although, as Dasgupta (1993, Chapter 2) points out, classical utilitarianism can be modified to incorporate constraints such that no utility is allowed to fall below some stipulated minimum. But this was never done to my knowledge.

[57] Though see the response in Harsanyi (1975).

[58] Pogge (1989) reviews this and other critiques of Rawls's principles of justice and provides a reinterpretation and (vigorous) defense of Rawls's original arguments.

version of maximin but from a different starting point, namely, the desire to equalize opportunities in society. This is premised on the view that poverty reflected exogenous circumstances facing individuals, as well as personal efforts. Severe empirical challenges remain in cleanly separating efforts from circumstances, but the conceptual distinction has bearing on thinking about antipoverty policy (as has long been recognized in policy debates reviewed below). In striving to equalize opportunities, we would not want to bring everyone down to a common but low level of opportunity. Instead, Roemer advocates that policy choices stemming from an "equal opportunity ethic" should maximize the welfare assigned to the worst off group, defined by a vector of exogeneous "circumstances"—those things that cannot be traced back to the choices made by the individual.[59]

Rawls opened the way to new nonutilitarian thinking on the conceptual foundations of antipoverty policy. This marked a return to the themes that emerged in the First Poverty Enlightenment, although these found more complete and rigorous formulations in the wake of the Second Poverty Enlightenment. Rather than being blamed solely on the bad behaviors of poor people, poverty came to be seen as stemming in large part from circumstances beyond their control, given circumstances of birth and market and governmental failures. This perspective gave promotional policies a deeper conceptual foundation. It was still granted that there was an important role for individual responsibility—that poverty did sometimes stem from bad choices. But this had ceased to be the dominant model. Careful opportunity-based formulations emerged in the writings of both philosophers (such as Cohen, 1989a and Arneson, 1989) and economists (including Roemer, 1998 and Fleurbaey, 2008).

So far the discussion has focused on the new philosophical and economic thinking of the Second Poverty Enlightenment. No less important to policy making were the new data, the new empirical research on those data, and the more popular writings and social movements of this time. We now turn to these.

22.6.2 The Rediscovery of Poverty in America

The industrialized world saw a boom in social spending in the second half of the twentieth century (Lindert, 2004).The new public attention to antipoverty policies is evident in the marked increase in references to "antipoverty," "poverty alleviation," and "redistribution" in the *Viewer* (Ravallion, 2011). References to "redistribution" peaked around 1980. "Redistribution of wealth" was often mentioned in the Great Depression, but use of this term declined during World War II and after until about 1960 when a new upsurge of interest emerged.

[59] This assumes that a unique vector exists, dominated by all others. Given that choices (efforts) vary, Roemer proposes to maximize the average welfare level of the worst off group, averaged across levels of effort.

The change in popular thinking was especially evident in the United States. In the wake of the civil rights movement (starting around 1955), the rediscovery of poverty in the midst of affluence was stimulated by important social commentaries, including Galbraith's (1958) *The Affluent Society* and Harrington's (1962) *The Other America*, both bestsellers at the time.[60] The success of Harrington's book was clearly a surprise; the first print-run was only 2500 copies, but by the mid-1990s it had sold 1.3 million copies.

Knowledge made this new awareness of poverty possible. The First Poverty Enlightenment lacked the theories and data that we take for granted today in measuring poverty, reckoning its costs, and informing public action. Nor was there much sign yet of the theories and movements that could represent the interests of poor people. That had changed by the 1950s. Authors like Harrington and Galbraith could formulate accessible knowledge-based arguments, including measurements from sample surveys. Many people were shocked in the early 1960s when the official calculations indicated that almost one-in-five Americans lived in poverty.

Although the type of quantification initiated by Booth and Rowntree 70 years earlier had been crucial, credibly reported qualitative observations in the media and popular books also had a huge influence, including on policy making at the highest levels. Many people were influenced by Harrington's efforts to "describe the faces behind the statistics" (p. 17). This was research aimed squarely at promoting change through knowledge. In an introduction to a 1993 reprint of *The Other America*, Howe (1993, p. xii) describes its central premise: ". . . that if only people *knew* the reality they would respond with indignation, that if people became aware of 'the invisible poor' they would act to eliminate this national scandal."

Galbraith and Harrington described a new "minority poverty" in America. A long period of poverty reduction had meant that the poor were now a minority, albeit a sizeable one. Although overall economic growth had allowed many of the "old poor" to move into the new middle class, others were left behind or thrown into poverty from which they could not escape. Widely held views about upward mobility and equality of opportunity in America also came into question based on empirical studies showing how much parental income and schooling affects the life chances of children (Duncan et al., 1972; Bowles and Gintis, 1976).

There were differences between Galbraith and Harrington in their understanding of this new poverty in America. Galbraith identified two reasons why so many of the old poor were unable to participate in the new opportunities. The first was physical or mental disability—what Galbraith called "case poverty"—while the second was that some were trapped in geographic pockets of poverty (his "insular poverty"). Although not rejecting these categories, Harrington argued that this was incomplete in that many of the minority poor had been negatively impacted by the same economic expansion that had benefited

[60] References to both books in the *Viewer* skyrocketed from the 1960s; the graph can be found here.

so many others. Significant economic change had created their poverty, and they were unable to recover. Here, Harrington is making an important point—that even propoor overall progress comes with losers as well as winners. And his description sounds a lot like the model of wealth dynamics in Section 22.2, whereby large negative shocks create persistent poverty, and recovery to get back on track is no small thing.

The political response in the United States included new social programs, notably under the Economic Opportunity Act of 1964, popularly known as the Johnson administration's War on Poverty (Sundquist, 1968). From early on, this policy effort was framed in nonutilitarian and nonwelfarist terms, especially emphasizing opportunities. The new programs included Head Start, which continues today (and is discussed further in Section 22.9). Data and knowledge to support the War on Poverty was provided by (among other bodies) a new national institute created in 1966, The Institute for Research on Poverty at the University of Wisconsin–Madison. This organization was charged with studying the causes of poverty in the United States and evaluating antipoverty programs.[61]

The War on Poverty was not, it seems, prompted by a mass shift in American public opinion; indeed, Heclo (1986) refers to US polls indicating that the public was evenly divided on whether welfare spending should increase. It seems that the political response was motivated by evidence and ideas, not attracting voters. Although causality is unclear, it is notable that the US poverty rate fell between the years 1960 and 1980 (Meyer and Sullivan, 2012).[62]

Similar to the First Poverty Enlightenment, a backlash emerged in due course. An influential counterattack came from Murray's (1984) *Losing Ground*. As was the case with the backlash from Malthus and others around the turn of the nineteenth century, concerns about adverse incentive effects on behavior returned to loom large, such as claims that welfare benefits to single mothers encouraged families to break up. However, as in the debates on the Old Poor Laws, rather little credible supportive evidence was presented, and evidence to the contrary could be cited (Ellwood and Summers, 1986; Hoynes, 1997). Yet reforms followed in the United States during the 1990s; 30 years after declaring a "War on Poverty," the American government declared a "War on Welfare."[63]

Although (again) the attribution to social policies alone can be questioned, it is notable that the decline in US income poverty rates up to about 1980 stalled, and even reversed after that. Also notable is that this came with a marked shift in the demographic

[61] A good history of the Institute can be found on their website.

[62] This is true for both incomes and consumptions; income poverty rates crept back up after 1980 though consumption-based measures continued to fall. On the choice between these measures, see Slesnick (2001).

[63] The latter term was used by Katz (1987); also see Albelda et al. (1996).

profile of US poverty, favoring the elderly. Indeed, the incidence of poverty continued to fall among the elderly in the United States after 1980, albeit at a slower rate. Lindert (2013) attributes this difference to a bias in US social spending in favor of the elderly over the young, in common with other rich countries.

In attempting to explain America's poverty amidst affluence, the ideas of a "culture of poverty" and an "underclass" that emerged in the 1960s were much debated. Echoing the debates of prior times (reviewed earlier), critics saw these ideas as blaming poor people for their poverty and ignoring more deep-rooted "structural" inequalities (Gans, 1995; O'Connor, 2002). In some versions of the "underclass" idea, such as in Wilson's (1987) *The Truly Disadvantaged*, a "culture of poverty" was seen to stem from structural inequalities and so was part of their explanation; echoing Harrington, Wilson emphasized macroeconomic factors, including structural changes in the economy, urban structural changes, and aggregate unemployment rates.

Although the debate continues about whether there is space for policy intervention aimed at changing culture,[64] looking back over 200 years, it is clear that there has been a significant shift in thinking about poverty, from primarily blaming poor people to identifying deeper factors beyond their control, yet amenable to public action. This new view did not deny personal responsibility or the scope for mistakes or seemingly irrational behaviors.[65] In due course, evidence also emerged that the stresses of poverty diminished cognitive ability (Mani et al. 2013), again clouding the issue of cause and effect. But the key point to emerge was that "bad choices" was a dangerously incomplete explanation of poverty. As Shipler (2005, p. 6) put it with reference to America's working poor: "Each person's life is the mixed product of bad choices and bad fortunes, of roads not taken and roads cut off by the accident of birth or circumstances."

22.6.3 Relative and Subjective Poverty

Before the Second Poverty Enlightenment, poverty was mainly seen in absolute terms.[66] This changed radically in many of the rich countries of the world from around 1960.[67] The Second Poverty Enlightenment saw a new concept of "relative poverty" in both America and Western Europe, where the idea attained widespread official acceptance. By this view, the definition of poverty was contingent on the average standard of living

[64] See, for example, Steinberg's (2011) comments on Small et al. (2010).

[65] Behavioral explanations of poverty have drawn some support from experiments suggesting that people do not always behave rationally, although the experiments are often open to other interpretations, notably about the nature of the optimizing behavior (Saint-Paul, 2011).

[66] By "absolute poverty," I mean a poverty line that is fixed in real terms over time.

[67] Doron (1990, p. 30) describes this change in the 1960s: "The reformers of the period, and certainly the radicals among them, rejected the absolute approach, which contents itself with guaranteeing a minimum of subsistence . . . The needs of men are not stable and absolute but relative and related to the circumstances of the society in a particular period of time."

in the society one was talking about and so could be expected to evolve with the average.[68] Fuchs (1967) appears to have been the first to propose the sharpest version of this idea: that the poverty line should simply be set at 50% of the current median income. For a reason that will soon be clear, I will call these "strongly relative measures."

Although all the debates in the United States that were reviewed earlier in this section were echoed across the Atlantic, this new idea of strongly relative poverty had more influence in Western Europe than in America, and it carried little or no weight in the developing world. In due course, the most widely used definition of poverty in Western Europe followed Fuchs's suggestion, with national poverty lines often set at a constant proportion of the current mean (or median). Eurostat (2005) has produced such relative poverty measures across European countries and over time, as has the influential Luxembourg Income Study (LIS), which started in the mid-1980s and uses a poverty line set at 40–60% of the median in its summary statistics at country level. An immediate implication of these measures is that, when all income levels rise by the same proportion, the measure of poverty remains unchanged.

There were antecedents to the idea of relative poverty in the First Poverty Enlightenment. As Himmelfarb (1984a,b) and others have observed, Adam Smith held a conception of poverty that was socially specific. In a famous passage in *The Wealth of Nations* (1776, Book 5, Chapter 2, Article 4), Smith pointed to the social role of a linen shirt in eighteenth-century Europe.[69] Smith, it seems, wanted the poverty line to be relevant to its context.

That is what we see across countries. The average poverty line rises from $1.25 a day for the poorest countries to $30 a day in the richest (Ravallion, 2012a). At around $13 per person per day, the official poverty line in the United States is far higher than the poverty lines found in poor countries (though below average for rich countries). However, strongly relative lines go further in that they are changing over time in direct proportion to the mean or median, that is, with an elasticity of unity. It is not clear that Smith had in mind such a definition of poverty. One might argue that the poverty line should be relative between countries but absolute for a given country. The official poverty line in the United States is still an absolute line over time (with fixed real value),[70] as are almost all poverty lines in developing countries (Ravallion, 2012a). Logically, however, a poverty line that is fixed in real terms cannot remain relevant to prevailing living standards

[68] The period also saw efforts to anchor poverty measures to governmental assistance thresholds; an early example was Abel-Smith and Townsend (1966) describing poverty in Britain. For further discussion of this and other approaches, see Atkinson (1991).

[69] In more recent times, a number of studies have also pointed to the social roles played by festivals, celebrations, and communal feasts; see, for example, Rao (2001), Banerjee and Duflo (2007), and Milanovic (2008).

[70] This has been set at three times the cost of an adequate diet, following Orshansky (1963). Supplementary measures have been introduced in recent years (Johnson and Timothy, 2012).

indefinitely in growing economies. Indeed, as Fuchs (1967) points out, the US poverty line in the 1930s was probably substantially lower in real terms than that of the 1960s.[71] Some gradient over time is clearly called for.

Although the idea of relative poverty goes back to the First Poverty Enlightenment (though largely dormant between the two Enlightenments), explicitly relative measures were a product of the Second Poverty Enlightenment. However, there has been much debate, and it continues today. Some observers have been concerned about unequal treatment of people at similar levels of real income. The advocates of relative poverty lines for rich countries would not presumably have been comfortable in applying the same idea in comparing poverty measures between the majority population and minorities within one country; indeed, the Second Poverty Enlightenment started to see a breakdown of past discriminatory practices in this respect. There were clearly (though rarely explicit) moral bounds to relativism. However, the case for relative poverty lines rested on the view that poverty must be seen as absolute in the space of welfare, whether defined in terms of utility or capabilities; as Sen (1983, p. 163) put it: "… an absolute approach in the space of capabilities translates into a relative approach in the space of commodities."

The more difficult issue was why the poverty line should be strongly relative, that is, proportional to the mean or median. If we consider more closely the two most common arguments made in favor of relativism, neither is compelling in this respect. The first argument concerns *social inclusion*. A linen shirt in eighteenth-century Europe is an example of what can be termed a "social inclusion need." The existence of such needs has been the primary justification given for the Western European relative poverty lines. However, the cost of that shirt will be roughly the same for the poorest person as the richest. More generally, the cost of social inclusion cannot be expected to go to zero in the limit, as mean income goes to zero, as implied by strongly relative lines. That would almost certainly understate the costs of social inclusion in poor countries.

The second argument made for the strongly relative measures is that they allowed for *relative deprivation*—that people care about their income relative to the mean or median of their country of residence.[72] However, this, too, is not so convincing on closer scrutiny. As long as we think that poverty is absolute in the space of welfare (or capabilities) one can only derive these strongly relative poverty measures if welfare *only* depends on relative income (own income relative to the median) (Ravallion, 2012a). In other words, one needs to assume that welfare does not depend on own income at given relative income. This must surely be considered a very strong assumption.

[71] Fuchs bases this claim on a necessarily rough calculation, asserting that if the 1960s standard in the United States was applied to the 1930s, then two-thirds of the US population would have been deemed poor as compared to President Roosevelt's estimate that "one-third of the nation" was poor in the 1930s.

[72] The sociologist Runciman (1966) was an influential advocate of this view.

None of this denies the welfare-relevance of social inclusion needs or relative deprivation. Arguably the case is now stronger than ever for incorporating relativist concerns in poverty measurement. Rather the issue is how best to do that. To allow for a (positive) minimum cost of social inclusion one requires what Ravallion and Chen (2011) dub "weakly relative measures."[73] These have the feature that the poverty line will not rise proportionately to the mean but with an elasticity less than unity for all finite mean incomes.[74] Consistent with the national poverty lines, Ravallion and Chen (2013) propose global poverty measures using a schedule of weakly relative poverty lines that contain the absolute lines (typical of poor countries) and relative lines (typical of rich ones) as the limiting cases.

Another strain of the new literature on poverty measurement emphasized the scope for calibrating welfare and poverty measures to subjective questions in surveys. These could take the form of a ladder (from "poor" to "rich" say),[75] or a more general question on satisfaction with life or happiness. Alternatively, the survey questions asked what income level corresponded to specific subjective welfare levels, following Van Praag (1968). A special case was the "minimum income question" that derived the monetary poverty line as the fixed point in the regression function relating personal subjective minima to actual incomes. In other words, the poverty line was drawn such that people with an income below it tended to think their income was inadequate for meeting their needs, while those above the line tended to think their own income was adequate. Alternatively, the poverty line could be identified as the fixed point of adequacy across multiple dimensions of welfare, following Pradhan and Ravallion (2000).[76]

22.6.4 The Rich World's Rediscovery of Global Poverty

A further surge of attention to poverty in the popular and scholarly literature in the late twentieth century stemmed from the Western public's increasing awareness of the existence of severe and widespread poverty in the developing world. Poverty and inequality in developing countries started to attract substantial mainstream scholarly attention in the West from the late 1960s.[77] GDP per capita was no longer seen as the sole metric for

[73] A weakly relative line was proposed earlier by Foster (1998). This was given by the weighted geometric mean of an absolute and a strongly relative line. Although this is also weakly relative, it has a constant elasticity, whereas the elasticity rises from zero to unity in the Ravallion and Chen (2011) proposal—consistent with the data on national lines.

[74] It can be argued that a globally relevant schedule of poverty lines should also have this property, and global measures following this approach are available in Ravallion and Chen (2013).

[75] These came to be known as Cantril ladders following Cantril (1965).

[76] For a critical survey of the various approaches found in this literature, see Ravallion (2014).

[77] Important contributions included Dandekar and Rath (1971), Adelman and Morris (1973), Chenery et al. (1974), Lipton (1977), World Bank (1980), Fields (1980), Kakwani (1980), Sen (1981a), Anand (1983), Bardhan (1984a), and Kanbur (1987).

judging success; for example, in his foreword to an overview by the World Bank of 25 years of development, the Bank's first Chief Economist, Hollis Chenery (Chenery (1977, p. v)), wrote that "... economic growth is a necessary but not sufficient condition for social progress and that more direct attention should be given to the welfare of the poorest groups."

For most of the developing world, poverty was "majority poverty"—in marked contrast to Galbraith's characterization of "minority poverty." Travel and visual media made it visible to those living in the West though it was already evident to almost everyone in the developing world. And poverty data were playing an important role in the post-Independence policy debates in some poor countries, including India, notably through its National Sample Surveys, which began in 1950.[78] As was the case with the poverty research by Booth and Rowntree in late nineteenth-century England, around 1990 many people were shocked to learn that there were about one billion people in the world living on less than $1 per day, at purchasing power parity (PPP) (Ravallion et al., 1991; World Bank, 1990)—an explicitly frugal line anchored to the national poverty lines found amongst the world's poorest countries.[79] Since 1990 there has been a massive expansion in survey data collection and availability and refinements to the methodology; the original estimates by Ravallion et al. (1991) used data for 22 countries, with one survey per country, while the latest estimates in Chen and Ravallion (2010) are based on survey data for 125 countries with more than six surveys per country on average. The efforts of country statistics offices—often with support from international agencies such as the UNDP, the World Bank, and the International Comparison Program—to collect household survey data and price data have provided the empirical foundation for domestic and international efforts to fight poverty since the 1980s. Public access to such data was crucial and gradually improved with help from efforts such as the World Bank's Living Standards Measurement Study (LSMS), which facilitates the collection of household-level survey data in developing countries, and the LIS, which facilitates access to harmonized micro data, though mostly for rich countries.

The World Bank's (1990) *World Development Report: Poverty* was influential in development policy circles, and soon after a "world free of poverty" became the Bank's overarching goal. A large body of empirical research on poverty followed in the 1990s, helped by a number of texts that provided useful expositions for practitioners of relevant theory and methods.[80] The UNDP's *Human Development Reports* began in 1990, and they have

[78] India was an early leader globally in the application of random sampling in economic and social statistics, notably through the Indian Statistical Institute, founded by the eminent statistician Prasanta Mahalanobis, which led in due course to India's National Sample Surveys, which are still used for measuring poverty in India.

[79] The $1 a day line was chosen as a typical poverty line for low-income countries. It was never exactly $1 a day, and the latest line (based on a much larger and more representative sample of national lines) is $1.25 a day at 2005 PPP (Ravallion et al., 2009).

[80] Examples include Ravallion (1994b), Sadoulet and de Janvry (1995), Deaton (1997), and Grosh and Glewwe (2000).

consistently argued for public action to promote basic health and education in developing countries. The importance to human development of combining poverty reduction with better access to basic services came to be appreciated (Anand and Ravallion, 1993). Sri Lanka's longstanding emphasis on basic health and education services had been shown to bring a large dividend in longevity and other human development indicators relative to countries at a similar level of average income (Sen, 1981b). The emphasis most East Asian countries have long given to broadly shared investments in human development also came to be recognized in the 1990s as a crucial element to their economic success, even though the role played by some other elements of the East Asian policy package remained contentious (Fishlow and Gwin, 1994; Rodrik, 1994; World Bank, 1993). It is clear that by the late twentieth century there had been a complete reversal in policy thinking about poverty, from the view 200 years earlier that human capital development for poor families was a waste of public resources to the view that it is an essential precondition for growth and development.

The period also saw a broadening of the range of policies under consideration, especially in the developing world. There was a new political will for antipoverty policy in many of the newly independent, postcolonial, states, although with mixed success. Policies for promoting economic growth came to be seen and judged by their efficacy in promoting (among other goals) poverty reduction (World Bank, 1990). (The next section will return to this point.) By the 1990s, it seems that nothing in the policy arena was off-limits in discussing impacts on poverty. This brought a new danger too. Without some degree of separability, allowing instruments to be tied to goals risked policy paralysis. But economic analysis and a measure of good sense could often be trusted to guide effective policy action, recognizing the tradeoffs. And the shift in focus from protection to promotion is also evident in the types of policies being tried within the subclass of direct interventions, as we will see in Section 22.9.

By the turn of the twenty-first century, a new optimism about the scope for global poverty reduction had emerged. The Millennium Development Goals (MDGs) were ratified in 2000 at the Millennium Assembly, a meeting of world leaders at the United Nations. The first MDG was to halve the developing world's 1990 "$1 a day" poverty rate by 2015. Sachs (2005b, p. 1) wrote a *New York Times* bestseller, *The End of Poverty*, outlining his personal vision of how "[o]ur generation can choose to end that extreme poverty by the year 2025." Some of this optimism was well founded in subsequent events. Using the $1.25-a-day poverty line based on 2005 prices, the first MDG was attained in 2010, a full 5 years ahead of the goal (Chen and Ravallion, 2013). Even so, that important achievement leaves over one billion people living in extreme poverty, as judged by the standards of the poorest countries. But continuing the success of the fight against extreme poverty begun in 2000 will lift one billion people out of extreme poverty by 2030 (Ravallion, 2013). Progress in reducing global relative poverty will be slower; today over 2.5 billion people remain poor by standards typical of the country they live in (Ravallion and Chen, 2013). However, national poverty elimination targets have emerged in many countries, both rich and poor. In 2010, the European Union adopted its Europe 2020 poverty reduction target to reduce by 25% the numbers of Europeans living below national poverty lines.

Some of the debates of 200 years ago survive today. For example, at the time of this writing, the US Congress was implementing substantial cuts to the Supplementary Nutrition Assistance Program (Food Stamps). During the relevant House of Representatives Committee Meeting, a Congressman was quoted as saying that "[w]hile it was a Christian duty to care for the poor and hungry, it was not the government's duty" (Fifield, 2013). One heard such claims often 200 years ago. The difference today is that the vast majority of people clearly do not agree.

Although there is continuing debate about the causes of poverty and policy prescriptions, modern writings are invariably based on the premise that poverty is something that can be greatly reduced and, indeed, eliminated with the right economic and social policies. By this view, poverty is in no small measure a *global public responsibility*, and governments and the economy are to be judged (in part at least) by the progress that is made against poverty.

22.7. THE IDEA OF A PROGRESSIVE MARKET ECONOMY

Until the late twentieth century, the prevailing view was one of skepticism that poor people would benefit much from economic growth in a capitalist economy. Well into the 1980s, it was common to hear in both popular and scholarly writings that economic growth was expected to largely bypass poor people in both rich and poor countries. Where did this skepticism come from, and was it justified?

By one view, poverty is likely to persist in a growing economy because poverty is relative (Section 22.6). Strictly, poverty could still be eliminated when using a strongly relative poverty line set at a constant proportion of the mean with sufficient redistribution in favor of poor people. Growth in the mean will not eliminate poverty without a change in relative distribution. However, the past explicit acceptance of poverty among economists and noneconomists alike does not appear to be the product of such a relativist view. In fact the latter is a modern idea, which appears to have emerged much later, in the 1970s (Section 22.6). Using absolute or weakly relative poverty measures, sufficient inequality-neutral growth will eliminate poverty.[81]

But growth was not expected to be inequality-neutral. Most classical and Marxist economic thinkers saw little hope that even a growing capitalist economy would deliver

[81] For any linear schedule of relative poverty lines as a function of the mean, and all standard measures of poverty, saying that there is a positive lower bound to the cost of social inclusion is essentially equivalent to saying that inequality-neutral growth in the mean will reduce the measure of poverty (Ravallion and Chen, 2011).

rapid poverty reduction or even *any* poverty reduction. Although Smith was optimistic about the potential for a progressive, poverty-reducing, market economy, the prominent classical economists who followed, including Malthus and Ricardo, were more pessimistic about the prospects for higher real wages and (hence) less poverty, suggesting that they anticipated rising inequality from a growing capitalist economy. As discussed in Section 22.5, demographic responses to rising wages were expected to play a key role in attenuating the poverty impact of growth. The socialist movement that emerged toward the middle of the nineteenth century shared the same pessimistic view about the prospects for poverty reduction but took it to be a damning criticism of capitalism. The thirst for profits to finance capital accumulation, combined with the large "reserve army" of unemployed, was seen as the constraint on rising real wage rates rather than population growth.

Distributional dynamics has long been a central theme of development economics. Poverty was a concern for the postcolonial governments of the newly independent countries, but the earliest policy-oriented discussions were pessimistic about the prospects of economic growth bringing much benefit to poor people. It was widely believed that growth in low-income countries was bound to be inequitable, and that view is still heard today.

A foundation for this view was provided by Kuznets (1955), and came to be known as the "Inverted U Hypothesis," whereby inequality first increases with economic growth in a poor country but falls after some critical income level is reached.[82] Although there are other theoretical models in the literature that can generate such a relationship, in the Kuznets formulation, the economy is assumed to comprise a low-mean, low-inequality rural sector and a high-mean, high-inequality urban sector, and growth is assumed to occur through the migration of workers from the former environment to the latter. This growth is assumed to entail that a representative "slice" of the rural distribution is transformed into a representative slice of the urban distribution, preserving distributions within each sector. An inverted U can then be derived linking certain indices of inequality and the population share of the urban sector (Anand and Kanbur, 1993; Robinson, 1976).

Some policy makers appear to have incorrectly inferred that this model also implied that economic growth in poor countries would bring little benefit to poor people. (This sometimes reflected a longstanding confusion between the ideas of "poverty" and "inequality" in development policy discussion.) It is easy to show that for all additive poverty measures, if poverty is initially higher in the rural sector, then aggregate poverty must fall under the Kuznets process of migration described above. Not for the last time,

[82] Also see Adelman and Morris (1973), Robinson (1976), Ahluwalia (1976), Ahluwalia et al. (1979), and Anand and Kanbur (1993).

thinking about how the overall development strategy might allow more rapid poverty reduction was led astray by misunderstandings of a theoretical model.

The economic history of today's rich countries has often been seen as a source of lessons for the developing world. Contrary to the expectations of both the nineteenth-century supporters and critics of capitalism, Britain's industrial revolution, which had started around 1760, almost certainly reduced poverty through rising real wage rates. But there was a long lag. Just how long depends on the position one takes in the debate about price indices. Clark's (2005) discussion of builders' real wage rates in England suggests that workers earned higher wages from about 1800, while Allen (2007, 2009) argues that the increase started closer to 1830. Either way the pessimists appear to have been right that for at least a few decades after the technical innovations, real wages did not increase.[83] Real wages in Britain did start to rise in the nineteenth century despite continuing population growth. Falling food prices in Europe due to refrigeration and lower freight transport costs also helped increase real wages later in that century (Williamson, 1998). And there is evidence that the gains in real wages for the working class from the mid-nineteenth century came hand-in-hand with improved nutrition.[84]

The lag in the real wage rate response to the industrial revolution is suggestive of the model by Lewis (1954) in which a surplus of labor in the rural economy keeps wages at a low level until that surplus is absorbed by the economy's modern (urban) sector, as this expands due to technical progress. Allen (2009) offers an alternative explanation whereby the extra demand for capital due to technical progress could only be met by savings from nonlabor income, under the assumption that workers were too poor to save. Then profits had to rise to finance the investments needed, and only when sufficient capital had accumulated did real wages rise. In short, high poverty rates had to persist for some time, despite growth, because poor people simply could not generate the savings required to support that growth. However, even a small amount of savings by each of a large number of workers could have helped finance capital accumulation provided that those savings could be mobilized. Financial underdevelopment may then be seen as a factor in the lag.

The empirical foundations for the expectation that inequality would inevitably rise in growing developing countries were not particularly secure at the time the Kuznets hypothesis was influential. There was not much data to draw on. A debate in the early 1970s on the distribution of the gains from economic growth in Brazil left an appetite for better survey data for measuring poverty and inequality.[85] As better evidence from household surveys accumulated, it was revealed that very few low-income countries have

[83] Also see Williamson (1985) and O'Rourke and Williamson (1997).

[84] See the Fogel et al. (1983) series on mean height of working class boys in London, which tracks quite closely Tucker's (1975) series on real wages of London artisans. However, Cinnirella (2008) puts the turning point (after which mean height rose) much later, around the mid-nineteenth century.

[85] Contributions to this debate were made by Fishlow (1972), Fields (1977), and Ahluwalia et al. (1980).

developed over time in a manner consistent with the Kuznets hypothesis, as is shown by Bruno et al. (1998) and Fields (2001). We have learned that growth in developing countries tends to be distribution-neutral on average, meaning that changes in inequality are roughly orthogonal to growth rates in the mean (Dollar and Kraay, 2002; Ferreira and Ravallion, 2009; Ravallion, 1995, 2001). Distribution-neutral growth implies that the changes in any standard measure of either absolute or weakly relative poverty will be negatively correlated with growth rates in the mean.

There is also evidence of inequality convergence, whereby inequality tends to increase in low inequality countries and decrease in high inequality countries (Bénabou, 1996; Ravallion, 2003). This is consistent with neoclassical growth theory, which shows that a fully competitive market economy contains forces for reducing inequality, as demonstrated by Stiglitz (1969) and Bénabou (1996). As Ravallion (2003) argues, the evidence we see of inequality convergence can also be explained by how economic policy convergence in the world during the 1990s interacted with prereform differences in the extent of inequality. To see why, suppose that reforming developing countries fall into two categories: those in which prereform controls on the economy were used to benefit the rich, keeping inequality artificially high (arguably the case in much of Latin America up to the 1980s), and those in which the controls had the opposite effect, keeping inequality low (as in Eastern Europe and Central Asia prior to the 1990s). Then liberalizing economic policy reforms may well entail sizable redistribution between the poor and the rich, but in opposite directions in the two types of countries.

The periods of global trade openness fostered some progress toward convergence of living standards across countries. Although much attention has been given to the current globalization period, Williamson (1998) argues that the prior period of globalization, 1870–1914, fostered economic expansion and convergence within the "Atlantic economy." This globalization almost certainly reduced poverty globally.

Post-Independence policies in most developing countries strived for economic growth, facilitated by government planning in relatively closed economies, although capabilities for effective implementation were often weak. India's Second and Third Plans, as well as many other planning documents, aimed for growth through accelerated capital accumulation and industrialization. These plans were influenced by classical economics and the Harrod–Domar equation, although here, too, policy makers misinterpreted the implications of the model.[86] The prioritization given to the capital-goods sector in India's Second Plan was directly influenced by a two-sector growth model in Mahalanobis (1953), although there were dissenters at the time (including Vakil and Brahmanand, 1956), and subsequent research in growth economics did not find any robust implication to justify this prioritization. As Lipton (1977) points out, the planners also ignored Adam Smith's warning that the food supply would constrain urban

[86] See the insightful discussion of the history of thought on economic growth in Ray (1998, Chapter 3).

growth in a closed economy. And poor people were financing the industrialization push, which typically depended on extracting a surplus from agriculture, which provided most of their incomes.[87] The plans were overly optimistic about rapid industrialization and about their potential to raise the demand for labor and so reduce poverty. And the industrialization push displaced other policies; for example, rural infrastructure (electrification and roads) took a back seat.

China's enormous progress against absolute poverty since around 1980, alongside rising inequality, might superficially be seen as testimony to the idea that the country has been in the rising segment of the Kuznets inverted U. However, here, too, the model just does not fit the facts. For one thing, inequality is lower in urban China than rural China, unlike the case assumed by Kuznets (1955), although this is not necessary for an inverted U; see Robinson (1976). More importantly, neither analytic decompositions of the changes in poverty nor regression-based decompositions suggest that the Kuznets process of growth through modern sector enlargement was the main driver of growth and poverty reduction in China (Ravallion and Chen, 2007). One must look elsewhere, notably to the initial agrarian reforms—including the massive land reform when the land of the collectives was assigned to individual farmers—and market liberalization more broadly, for an explanation of China's rapid poverty reduction in the 1980s.[88] Manufacturing growth came to play an important role later though that success was based in part on favorable initial conditions, notably the legacy of investments in human development, including in rural areas. Unlike many developing countries, there was a large literate rural population to draw on as the workforce for China's labor-intensive modern sector enlargement.

In thinking about policies for fighting poverty, the role played by the rural sector has been much debated. The sequence in China was roughly right: In the reform period from 1978, initial attention was given to the rural sector, and agrarian reforms to restore farmer incentives (in land allocation and prices) were crucial to ensuring a sustainably propoor development path, as had been the case elsewhere in East Asia.[89] Few other countries got the sequence right, and China's experience contains an important lesson for Africa today (Ravallion, 2009).

There were efforts to reprioritize development policy in the 1970s and 1980s. World Bank President Robert McNamara's 1973 "Nairobi speech" signaled such an effort from the international development institutions. In development thinking, "urban bias" was increasingly recognized as bad for growth as well as for poverty reduction, though it reflected political structures in much of the developing world (Lipton, 1968, 1977). However, the

[87] Even now, three-quarters of the developing world's poor live in rural areas (Ravallion et al., 2007).

[88] Similarly, in Taiwan and South Korea, the initial conditions for more propoor growth were laid by radical redistributive land reform, which led to productive and dynamic owner-farmed smallholdings.

[89] Given that the rest of the economy was growing rapidly, China could delay reforms to its state-owned enterprises (SOEs). Indeed, it was not until the late 1990s (20 years after the agrarian reforms began) that China started reforming its SOEs. Some observers have suggested that this should have happened sooner.

temptation to industrialize rapidly—"run before you have walked"—was strong. Combined with huge inequities in access to finance and human development, the subsequent growth paths were disappointing, both in growth and (especially) poverty reduction.

The debt crises of the 1980s brought a wave of structural adjustment programs supported by the international financial institutions (IFIs) that attempted to restore macroeconomic balances and promote economic growth. Given that the World Bank had produced *Redistribution with Growth* 10 years earlier (Chenery et al., 1974), it is surprising that its own adjustment programs in the early and mid-1980s gave little serious attention to the impacts on poor people though this neglect was consistent with the broader 1980s backlash in the Anglo-Saxon world against the distributional focus of the 1960s and 1970s. The Bank and Fund programs were much criticized for their neglect of distributional impacts, and the criticisms stuck. A progressive recovery in thinking within the IFIs was underway by the late 1980s, and add-on programs to "compensate the losers from adjustment" were soon common. Today, it is widely recognized that poverty and inequality mitigation has to be designed into economy-wide reform programs from the outset.

By the turn of the twenty-first century, enough evidence had accumulated for economists to be confident that higher growth rates tended to yield more rapid rates of absolute poverty reduction.[90] A more poverty-reducing process of global economic growth emerged after 2000, and not just because of China's growth. The trend rate of decline in the "$1.25-per-day" poverty rate for the developing world outside China rose from 0.4% points per year from 1980 to 2000 to 1.0% points per year after that period (Ravallion, 2013).

The poverty impact of a given rate of growth depends in part on the initial distribution.[91] Intuitively, when inequality is high, poor people will tend to have a lower share of the gains from growth. Ravallion (1997a, 2007) confirmed this using household survey data over time.[92] Easterly (2009) conjectured that the initial poverty rate is likely to be the better predictor of the elasticity than initial inequality though no evidence was provided. Ravallion (2012b) provided that evidence, and it compellingly shows that it is not high initial inequality that impedes the pace of poverty reduction at a given rate of growth, but high poverty.

Saying that growth typically reduces poverty does not, of course, mean that any growth-promoting policy will do so or that everyone will benefit. That depends on

[90] See Ravallion (1995, 2001, 2007), Fields (2001), Dollar and Kraay (2002), Kraay (2006), and World Bank (1990, 2000). Also see the review of the arguments and evidence on this point in Ferreira and Ravallion (2009).

[91] See Ravallion (1997a, 2007, 2012b), World Bank (2000, 2006), Bourguignon (2003), and Lopez and Servén (2006).

[92] Ravallion (1997a) did not find that the elasticity of poverty to growth varied systematically with the mean, although if incomes are log-normally distributed, then such a variation is implied theoretically (Bourguignon, 2003; Lopez and Servén, 2006).

the distribution—horizontally as well as vertically—of the gains and losses from that policy. There may be vertical inequalities—between people at different levels of mean income—generated in the process that mitigate the gains to poor people from growth. And there can be horizontal inequities, whereby people at the same initial levels of income fare very differently, and some poor people may well lose from a policy that reduces poverty in the aggregate. (Recall that Harrington (1962) emphasized this point in describing the new "minority poverty" in the "other America.")

This point has been clearest in the literature on external trade and poverty. A number of studies have found support for the view that trade openness—typically measured by trade volume as a share of GDP—promotes economic growth.[93] It is unclear that trade volume can be treated as exogenous in these cross-country regressions; higher trade volume may be a response to growth rather than a cause. The policy implications are also unclear since trade volume is not a policy variable; see the discussion in Rodrik (1994) and Rodriguez and Rodrik (2001). But, putting this issue to one side, what about the distributional effects? A number of studies have combined survey-based measures of income inequality at country-level with data on trade and other control variables to assess the distributional impacts of trade openness, as reviewed in Winters et al. (2004). The evidence is mixed. Dollar and Kraay (2004) find little or no effect of trade volume on inequality. Other studies have reported adverse effects. Lundberg and Squire (2003) find evidence that higher trade volume tends to increase inequality. On balance, Ravallion (2006) reports little or no correlation between greater trade openness and the pace of poverty reduction in developing countries.

However, there can be winners and losers at all levels of living, even when a standard measure of inequality or poverty is unchanged. There are many sources of heterogeneity, yielding horizontal impacts of reform. Geographic disparities in access to human and physical infrastructure affect prospects for participating in the opportunities created by greater openness to external trade. Differences in household demographic composition influence consumption behavior and hence the welfare impact of the shifts in relative prices associated with trade openness. Ravallion (2006) reports on two case studies of this heterogeneity in the welfare impacts of liberalizing trade reform, for China and for Morocco. The results indicate a sizable, and at least partly explicable, variance in impacts across households with different characteristics—differences that influenced their net trading positions in the relevant markets.

Where does all this leave us? The antitrade policies (on quotas, tariffs, and exchange rates) of the post-Independence development policy regimes were unlikely to bring

[93] In a metastudy of all the cross-country growth regressions with an average of seven regressors (chosen from 67 candidates drawn from the literature on cross-country growth regressions), Sala-I-Martin et al. (2004) report that trade volume is a significant factor in two-thirds of the regressions, though it is not among their subset of 18 robust predictors of economic growth.

much benefit to poor people, the bulk of whom produced tradable goods from primarily nontradable inputs. Although this remains a plausible generalization, there is likely to be considerable heterogeneity across countries in such effects, and one might be skeptical of basing policy advice for any specific country on generalizations from either standard Stolper–Samuelson arguments or cross-country regressions (Ravallion, 2006). For example, some studies have found evidence that higher trade volume increases inequality in poor countries but that the reverse holds true at a higher mean income (Milanovic, 2005; Ravallion, 2001). The macro perspective, focusing on impacts on an aggregate measure of poverty or inequality, hides potentially important horizontal impacts with implications for other areas of policy, notably social protection efforts that may well be needed to complement the growth-promoting reforms. (Section 22.9 discusses these policies further.)

Trade policies have also played a role in social protection, though this, too, has been much debated. Governments of food-exporting but famine-affected areas have often implemented food export bans in the hope of protecting vulnerable citizens. Classical economists were influential in arguing against such policies in favor of free trade. For example, Aykroyd (1974) describes how the Governor of Bombay in the early nineteenth century quoted Smith's *The Wealth of Nations* when defending his policy stance against any form of trade intervention during the famines that afflicted the region. Various "Famine Commissions" set up by the British Raj argued against the trade interventions that were being called for to help protect vulnerable populations. Similarly, Woodham-Smith (1962) describes the influence that Smith and other classical economists had on British policy responses to the severe famines in Ireland in the mid-nineteenth century. In modern times, free trade has been advocated as a means of stabilizing domestic food consumption in the presence of output shocks (World Bank, 1986). Other economists have been less supportive. Sen (1981a) and Ravallion (1987) pointed to the possibility that real income declines in the famine-affected areas can generate food exports while people starve.[94] Regulated trade through taxes or even export bans may then be a defensible policy response to help vulnerable groups relative to feasible alternatives (Ravallion, 1997b).

Critics of trade intervention for the purpose of protection from external price shocks (such as in the period from 2007 to 2011) have pointed out that such a policy can exacerbate the problem of price volatility (Martin and Anderson, 2012). However, in the absence of better options for aggregate intertemporal smoothing, the optimal nontrade protection policy would entail transfers between net food producers and net consumers, to coinsure. And this, too, would exacerbate the price volatility, as shown by Do et al.

[94] The analysis of the time series data for famines in British India in Ravallion (1987) indicated that the aggregate income effects were not strong enough to undermine the consumption-stabilizing effects of unrestricted trade.

(2013). So one cannot simply argue that external trade intervention is an inferior form of social protection; any such protection would have a similar feature. Trade interventions will probably entail some price distortions, which must be evaluated against the distortions generated by alternative schemes. There are situations in which trade insulation dominates feasible options for protection (Do et al., 2013).

The key point here is to avoid sweeping generalizations about policies. To take another example (possibly even more contentious than trade policy) consider active industrial policies—the effort to encourage selected promising sectors or firms using tariffs, subsidies, or tax breaks.[95] Advocates point to the successes of some East Asian countries with these policies, though sometimes downplaying the failures of other countries with similar policies. Instead of arguing for or against such policies in the abstract, the focus should be on understanding under what conditions these, or other interventions, work.

Possibly any country will have a good chance of success with a reasonably wide range of policies in a context of macroeconomic stability and a capable public administration that can pragmatically choose sensible interventions and minimize the damage from mistaken ones. But will that be enough? The next section turns to another set of potentially important initial conditions related to the distribution of wealth and income.

22.8. THE FINAL BLOW TO THE IDEA OF THE UTILITY OF POVERTY?

A strain of thought dating back to the mercantilists has essentially argued that, whatever moral position one takes about poverty, a more unequal initial distribution of income allows a higher long-run mean income for any given initial mean. Since higher inequality at a given initial mean almost certainly entails higher poverty (by any standard measure) this amounts to an instrumental excuse for higher poverty now. In other words, by this view, one need not worry about poverty today as it will come with higher growth and (hence) less poverty in the future.

The precise form of this argument evolved over time, although incentives always played a role. Mercantilists worried about adverse effects of higher wages on work effort and export competitiveness. Later arguments switched to the idea that aggregate savings constrained growth. By this view, in a fully employed (closed) economy, capital accumulation was constrained by aggregate domestic savings, and saving is something rich people naturally do more of than poor people. Thus—the argument went—efforts to redistribute income in favor of the poor risked retarding growth and (hence) had ambiguous implications for poverty reduction.

[95] A good review of this class of policies and the debate surrounding them can be found in Harrison and Rodríguez-Clare (2010). Supportive discussions can be found in Rodrik (2004) and Lin (2012); a more critical perspective can be found in Pack and Saggi (2006).

The neoclassical theory of economic growth, as represented by the Solow (1956) model, was interpreted by some observers as implying that there was an automatic self-correcting process whereby a high initial level of poverty would eventually be reduced by economic growth. By this argument, countries starting out with a low mean income (and hence high absolute poverty rate) would tend to have a higher marginal product of capital (given that they had so much less capital per worker and that there are diminishing returns), which would entail a higher rate of economic growth when compared to growing high income countries with a similar rate of investment. And so the initially poorer country would eventually catch up. This was strictly a process of dynamic transition, not a model for explaining differences in the steady-state level of income. However, with suitable controls for the latter, a body of empirical work confirmed the prediction of conditional convergence, following an influential early contribution by Barro and Sala-i-Martin (1992).

Because the Solow model is an aggregate model, with no heterogeneity, it was questionable to use it to argue that poverty would be self-correcting. There was no inequality in this model.[96] And, even in his aggregate model, Solow was well aware of the potential for a "poverty trap" (though he did not use that term). Indeed, the original (1956) paper outlined one possible trap, arising from assumed nonlinearities in how population growth rates depend on mean income, with population growth falling at low incomes but rising with higher incomes, then tapering off at higher incomes. A country in a stable equilibrium but at low income would then need a large gain in capital per worker to escape the trap and move to a sustainably positive growth path.

The twentieth century saw another set of ideas, which challenged the "utility of poverty." (Recall that there was an early hint of this challenge in Marshall (1890)). It appears to have been long understood that rich people saved a greater share of income than poor people, who were often assumed to save nothing (as in the models of Kalecki, 1942 and Kaldor, 1955). It would then have been only a small step to the conclusion that a higher poverty rate at a given mean income would yield lower aggregate savings and (hence) a lower growth rate in any economy for which aggregate savings constrained growth. But that conclusion was never drawn to my knowledge. It was, however, understood at least starting in the 1930s that the same property of the savings function implied a growth-equity tradeoff, whereby higher inequality would generate higher savings and (hence) higher growth. Keynes (1936, Chapter 24) questioned the existence of such a tradeoff. His interpretation of the causes of unemployment predicted that it was lack

[96] There was much debate around this time concerning the assumption of an aggregate neoclassical production function, such as in the Solow model, which ignored the heterogeneity of capital. Defenders of that assumption argued that it was an analytically useful simplifying assumption, albeit an assumption that became a workhorse of modern macroeconomics. There is an insightful discussion of this debate in Bliss (1975, Chapter 8).

of consumption that prevented full employment, and so a higher share of national income in the command of poor people would promote growth until full employment was reached.

In the 1990s, a new set of ideas emerged that seriously questioned the instrumental case for poverty *and* inequality even in a fully employed economy. By this view, poor and/or unequal societies stifled investment, invention, and reform.[97] These ideas opened up a new window to the potential role of antipoverty policies in economic development.

One argument about why poverty would self-perpetuate in the absence of effective policies related to the idea that poverty would foster a high rate of population growth which would (in turn) entail lower growth. The last step in this argument is an implication of the Solow model discussed above. In that model, a higher rate of growth of the labor force dilutes the capital stock. A higher rate of population growth thus acts in a similar way to a higher rate of depreciation in lowering the steady-state level of capital per worker and (hence) mean income.[98] But what about the first step? The modern version of this argument emphasizes the role played by inequality. An undeniably important dimension of inequality in the world is that people living in poorer families tend to be less healthy and to die sooner. This and other factors—including a dependence on children for old-age support and inequalities in maternal education—play a key role in generating another socioeconomic gradient: fertility rates tend to be higher in poor families. On balance, the natural rate of population growth also tends to be higher for the poor. Thus, we can expect lower rates of progress against poverty in countries with higher population growth rates, and there is some supportive evidence for this view.[99]

An influential strain of thought in the late twentieth-century literature also pointed to the implications of borrowing constraints associated with asymmetric information and the inability to write binding enforceable contracts. Credit market failure leaves unexploited opportunities for investment in physical and human capital, and there is assumed to be a diminishing marginal product of capital. (This idea can also be extended to embrace technical innovation, assuming that everyone gets new ideas, but that the poor are more constrained in developing these ideas.) Then higher current inequality implies lower future mean wealth at a given value of current mean wealth.[100]

[97] See Loury (1981), Banerjee and Newman (1993), Perotti (1996), Hoff (1996), Aghion et al. (1999a,b), Bardhan et al. (2000), Ghatak and Jiang (2002), Banerjee and Duflo (2003), Azariadis (2006), and World Bank (2006, Chapter 5). Voitchovsky (2009) provides a survey of the arguments and evidence for how the initial level of inequality influences the subsequent growth rate.

[98] Evidence of an adverse effect of population growth on GDP per capita growth can be found in Kelley and Schmidt (1995, 2001) and Williamson (2001).

[99] Evidence can be found in Eastwood and Lipton (1999, 2001), who regressed changes over time in poverty measures for a cross section of countries on the fertility rate (with various controls) and found an adverse demographic effect on poverty. Using time series data for India, Datt and Ravallion (1998) found evidence that higher rates of population growth increased poverty.

[100] Models with such features include Loury (1981), Galor and Zeira (1993), Bénabou (1996), Aghion and Bolton (1997), and Banerjee and Duflo (2003).

The model outlined at the beginning of this chapter illustrates this point well in the special case in which the distribution of wealth (given production technologies) is such that the threshold is not binding ($w_t > k^{\min}$ for everyone). Mean future wealth in a growing economy is then a weakly quasi-concave function of the distribution of current wealth. By standard properties of such functions, a mean-preserving increase in wealth inequality will entail lower mean wealth in the future, that is, a lower growth rate (Banerjee and Duflo, 2003). This is no longer true in general when the threshold is binding. Then there will exist increases in inequality embracing the lower end of the wealth distribution (below $k^{\min}$) that can increase the growth rate of wealth. Thus, the type of model illustrated by Figure 22.1 has ambiguous implications for how much an exogenous reduction in inequality will promote overall growth. That depends crucially on precisely *where* in the distribution the reduction in inequality occurs.

Borrowing constraints is not the only way that inequality can matter to growth. Other models have also been proposed, implying that high inequality leads democratic governments to implement distortionary redistributive policies, as in the model of Alesina and Rodrik (1994). Another class of models is based on the idea that high inequality restricts efficiency-enhancing cooperation, such that key public goods are underprovided, or desirable economic and political reforms are blocked.[101] Rajan (2009a,b) provides an interesting analysis of how the two main types of economic reforms that are widely seen as key to poverty reduction, namely, making markets more competitive and expanding access to education, can be blocked in a democracy in which three classes—the rich oligopolists who benefit from market distortions, an educated middle class, and the uneducated poor supplying unskilled labor—strive to preserve their rents in the status quo. The model helps us understand the observations of Weiner (1991) and others about India's relative lack of progress in attaining mass literacy.

A new interpretation of the long-term impacts of colonialism has identified adverse effects of initial inequality on policies and institutions; Engerman and Sokoloff (2006) provide an overview. The essence of this argument is that the geographic patterns of colonialism (notably between North and South America) implanted greater initial inequality and population heterogeneity in some colonies than others. The main colonial origin of inequality is seen to have been the creation of European enclaves in the colonies that were greatly advantaged over the natives. The more unequal colonies had a harder time developing promotional antipoverty policies (such as mass schooling) that were favorable to both long-term growth and poverty reduction.

But is it inequality that matters to growth and poverty reduction or something else, such as poverty, the size of the middle class, or the extent of polarization? Inequality is obviously not the same thing as poverty; inequality can be reduced without a lower poverty measure by redistributing income among the nonpoor, and poverty can be reduced

[101] Arguments along these lines include Bardhan et al. (2000), Banerjee and Iyer (2005), Acemoglu and Robinson (2006), Rajan (2009a,b), and Stiglitz (2012).

without lower inequality. (Similarly, efforts to help the middle class may do little to relieve current poverty.) In fact, there is another implication of credit market failures that has received less attention until recently. Although the literature has emphasized that higher inequality in such an economy implies lower growth, so, too, does higher current wealth poverty for a given mean wealth.[102] Again, the point can be illustrated using the model outlined in Section 22.2. Plainly, a larger density of people near the zero wealth equilibrium will entail lower subsequent growth. What if the threshold is not binding? It is assumed that the poverty line does not exceed $k^{\star}/(\lambda+1)$, and we can let H_t^* denote the poverty rate (headcount index) at this maximum poverty line. Now consider the growth effect of a mean-preserving increase in the poverty rate. I assume that H_t^* increases and that no individual with wealth less than $k^{\star}/(\lambda+1)$ becomes better off. If this holds true, then we can say that poverty is unambiguously higher. Then the credit constraint implies that unambiguously higher poverty incidence—defined by any poverty line up to the minimum level of initial wealth needed to not be liquidity constrained—yields lower growth at a given level of mean current wealth. As this point does not appear to have been made in the literature, the Appendix demonstrates the point more formally.

This theory implies an aggregate efficiency cost of a high incidence of poverty. But note that the theoretical prediction concerns the level of poverty at a *given* initial value of mean wealth. Without controlling for the initial mean, the sign of the effect of higher poverty on growth is ambiguous (see the Appendix). Two opposing effects can be identified. The first is the conditional convergence property described above, whereby countries with a lower initial mean (and hence higher initial poverty) tend to have higher subsequent growth in a neoclassical growth model. Against this, there is an adverse distributional effect of higher poverty. Which effect dominates is an empirical question, which we will return to later in the chapter.

Credit market imperfections are not the only argument suggesting that poverty is a relevant parameter of the initial distribution. Lopez and Servén (2009) introduce a subsistence consumption requirement into the utility function in the model by Aghion et al. (1999a) and show that higher poverty incidence (failure to meet the subsistence requirement) implies lower growth. Another example can be found in the theories that have postulated impatience for consumption (high time preference rates possibly associated with low life expectancy) and hence low savings and investment rates by poor people (see, for example, Azariadis, 2006). Here, too, although the theoretical literature has focused on initial inequality, it can also be argued that a higher initial incidence of poverty means a higher proportion of impatient consumers and hence lower growth.

The potential inefficiency of poverty is starkly obvious when one considers how work productivity is likely to be affected by past nutritional intakes, as these determine the stock

[102] Ravallion (2001) argued intuitively that poverty retards growth when there are credit market failures.

of human capital.[103] As noted in Section 22.2, only when nutritional intake is high enough will it be possible to do any work, but diminishing returns to work will set in later; see the model in Dasgupta and Ray (1986). Poverty's effects on the nutrition of young children in poor families are also of special concern. A sizable body of research suggests that poor nutrition (both food energy intakes and micronutrients) in the early years of life retards children's growth, cognitive and learning abilities, schooling attainment, work productivity, and likely earnings in adulthood.[104] The health environment also matters. Chronic undernutrition in children can stem from either low nutritional intake or low nutritional absorption due to constant fecal–oral contamination,[105] such as due to the lack of clean drinking water. This can mean that direct nutritional supplementation does little or nothing to improve children's nutritional status (such as measured by stunting) until the health environment improves.[106] This type of argument can be broadened to include other aspects of child development that have lasting impacts on learning ability and earnings as an adult (Cunha and Heckman, 2007). And the handicap of poverty can emerge in the prenatal period. Maternal and prenatal conditions are now also thought to matter to child development and (hence) economic outcomes later in life (Currie, 2011; Dasgupta, 2011). By implication, having a larger share of the population who were born in and grow up in poverty (including living in poor health environments) will have a lasting negative impact on an economy's aggregate output. Poverty will perpetuate.

In another strain of thinking about how poverty can perpetuate, Mani et al. (2013) present evidence from both experimental and observational studies, suggesting that poverty reduces cognitive ability. The evidence is consistent with the view that, given that human cognitive capacity is physically limited, the concerns generated by poverty crowd out thinking about other things relevant to personal economic advancement.

There are also theoretical arguments involving market and institutional development, although this is not a topic that has received as much attention in the literature. Although past theories have often believed credit market failures to be exogenous, poverty may well be a deeper causative factor in financial development (as well as an outcome of the lack of financial development). For example, given a fixed cost of lending (both for each loan and for setting up the lending institution), liquidity constraints can emerge as the norm in very poor societies.

Some of the theoretical literature has also pointed to the possibilities for multiple equilibria associated with a nonconvexity in the production possibility set, as in Figure 22.1. As noted already, in poor countries, the nutritional requirements for work

[103] Strauss and Thomas (1998) review evidence of this relationship. A useful overview of the biomedical arguments and evidence can be found in Dasgupta (2011).

[104] For useful overviews of the evidence, see Alderman et al. (2006), Benton (2010), and Currie (2011).

[105] This is known as environmental enteropathy (see, for example, Korpe and Petri, 2012).

[106] Kinsey (2013) identifies this as one possible reason why the incidence of chronic undernutrition has not fallen in his panel data for Zimbabwe.

can readily generate such nonlinearity in the dynamics, as argued by Dasgupta (1997). Such a model predicts that a large exogenous income gain may be needed to attain a permanently higher income and that seemingly similar aggregate shocks can have dissimilar outcomes; growth models with such features are also discussed in Day (1992) and Azariadis (1996, 2006) among others. Sachs (2005a,b) has invoked such models to argue that a large expansion of development aid would be needed to ensure a permanently higher average income in currently poor countries.

Some of the empirical literature on economic growth has found that higher initial inequality impedes growth.[107] And the effect is quantitatively large, as well as statistically significant. Consider the two most recent published studies at the time of this writing. Herzer and Vollmer (2012) find that a 1% point increase in the Gini index results in a decrease in long-term mean income of 0.013%; when normalized by standard deviations, this is about half the growth impact of the investment share. Berg et al. (2012) also find that more unequal countries tend to have less sustained spells of growth, and this effect is also quite large; a 1% point higher Gini index is associated with a decline in the length of the growth spell of 11–15%.

Not all the evidence has been supportive.[108] The main reason why some studies have been less supportive appears to be that they have allowed for additive country-level fixed effects in growth rates. This specification addresses the problem of time-invariant latent heterogeneity in growth rates. However, it may well have little power to detect the true relationships given that the changes over time in growth rates will almost certainly have a low signal-to-noise ratio. Simulation studies have found that the coefficients on growth determinants are heavily biased toward zero in fixed-effects growth regressions (Hauk and Wacziarg, 2009).

There are a number of remaining issues in this literature. The bulk of the literature has used consumption or income inequality measures. Theoretical arguments based on borrowing constraints point to the importance of asset inequality, not income inequality *per se*. There is evidence of adverse effects of asset inequality on growth.[109]

The aspect of initial distribution that has received almost all the attention in the empirical literature is inequality, as typically measured by the Gini index of (relative) inequality. The popularity of the Gini index appears to owe more to its availability in secondary data compilations on income and consumption inequality measures than to any intrinsic relevance to the economic arguments.[110] However, as Lopez and Servén

[107] See Alesina and Rodrik (1994), Rodrik (1994), Persson and Tabellini (1994), Birdsall et al. (1995), Clarke (1995), Perotti (1996), Deininger and Squire (1998), Knowles (2005), Voitchovsky (2005), Herzer and Vollmer (2012), and Berg et al. (2012).

[108] See Li and Zou (1998), Barro (2000), and Forbes (2000).

[109] See Rodrik (1994), Birdsall and Londono (1997), and Deininger and Olinto (2000), all using cross-country data, and Ravallion (1998), using regional data for China.

[110] The compilation of Gini indices from secondary sources (not using consistent assumptions) in Deininger and Squire (1996) led to almost all the tests in the literature since that paper was published.

(2009) observe, the significance of the Gini index in past studies may reflect an omitted variable bias, given that one expects that inequality will be highly correlated with poverty at a given mean.

There are also issues about the relevant control variables when studying the effect of initial distribution on growth. The specification choices in past work testing for effects of initial distribution have lacked clear justification in terms of the theories predicting such effects. Consider three popular predictors of growth, namely, human development, the investment share, and financial development. Of the first predictor, basic schooling and health attainments (often significant in growth regressions) are arguably one of the channels linking initial distribution to growth. Indeed, that is the link in the original papers of Loury (1981) and Galor and Zeira (1993).[111] The second predictor, one of the most robust predictors of growth rates, is the share of investment in GDP (Levine and Renelt, 1992); yet, arguably one of the main channels through which distribution affects growth is via aggregate investment, and this investment is one of the channels identified in the theoretical literature. Finally, consider private credit (as a share of GDP), which has been used as a measure of "financial sector development" in explaining growth and poverty reduction (Beck et al., 2000, 2007). The theories discussed above based on borrowing constraints suggest that the aggregate flow of credit in the economy depends on the initial distribution.

Although the theories and evidence reviewed above point to inequality and/or poverty as the relevant parameters of the initial distribution, yet another strain of the literature has pointed to various reasons why the size of a country's *middle class* can matter to the fortunes of those not (yet) so lucky to be middle class. It has been argued that a larger middle class promotes economic growth by fostering entrepreneurship, shifting the composition of consumer demand, and making it more politically feasible to attain policy reforms and institutional changes conducive to growth.[112] This has been an issue in India, where, since the 1970s, it has been argued that "inequality" constrained the growth of the manufacturing sector by limiting the size of the domestic market for consumer goods; see, for example, the discussion in Bardhan (1984b, Chapter 4). Here, too, it can be argued that it was not inequality *per se* that was the culprit but the relatively small middle class, or (more or less equivalently) the extent of absolute poverty that generated the domestic demand constraint in a relatively closed economy. The argument has been heard less in the more open economies. However, the Indian middle class has also been seen to promote reform (Sridharan, 2004). Using cross-country regressions, Easterly (2001) finds

[111] More recently, Gutiérrez and Tanaka (2009) have shown how high initial inequality in a developing country can yield a political economy equilibrium in which there is little or no public investment in basic schooling; the poorest families send their kids to work, and the richest turn to private schooling.

[112] Analyses of the role of the middle class in promoting entrepreneurship and growth include Acemoglu and Zilibotti (1997) and Doepke and Zilibotti (2005). Middle class demand for higher quality goods plays a role in the model of Murphy et al. (1989). Birdsall et al. (2000) conjecture that support from the middle class is crucial to reform.

that a larger income share controlled by the middle three quintiles is a significant predictor of rates of economic growth.

So we have three main contenders for the distributional parameter most relevant to growth: inequality, poverty, and the size of the middle class. The fact that very few encompassing tests are found in the literature and that these different measures of distribution are not independent, leaves one in doubt about what aspect of distribution really matters. As already noted, when the initial value of mean income is included in a growth regression alongside initial inequality, but initial poverty is an excluded but still relevant variable, the inequality measure may pick up the effect of poverty rather than inequality *per se*. Similarly, the main way the middle class expands in a developing country is almost certainly through poverty reduction, so it is unclear whether it is a high incidence of poverty or a small middle class that impedes growth. Similarly, a relative concept of the "middle class," such as the income share of middle quintiles, will probably be highly correlated with a relative inequality measure, clouding the interpretation.

Possibly, the strongest evidence to date to support the view that it is poverty not inequality *per se* that impedes growth in developing countries comes from an observation made by Ravallion (2012b), namely, that we see convergence in average living standards among developing countries and greater progress against poverty in faster growing economies, yet we do not see poverty convergence; the poorest countries are not enjoying higher proportionate rates of poverty reduction. Ravallion resolves this paradox by arguing that a high initial incidence of poverty, at a given initial mean, impedes subsequent growth (this theory is compatible with a number of the theories outlined above). This is shown to be consistent with data for almost 100 developing countries, which reveal an adverse effect on consumption growth of high initial poverty incidence at a given initial mean. Ravallion finds that high poverty at a given initial mean matters more than inequality or measures of the middle class or polarization. Also, starting with a high incidence of poverty limits progress against poverty at any given growth rate. For many poor countries, the growth advantage of starting out with a low mean is lost due to their high poverty rates. That does not, however, imply that any antipoverty policy will promote growth. That will depend on many factors, as discussed in the next section.

The arguments summarized above about why poverty can bring lasting efficiency costs do not require the existence of a poverty trap. However, when a poverty trap is present, the cost of poverty can rise greatly. So it is important to ask whether such traps have economic significance. On *a priori* grounds, it is highly plausible that threshold effects exist. Biology alone makes this plausible; unless one can support the nutritional needs of the body at rest, it will be impossible to do any work. Whether this is of economic significance in practice (even in poor economies) is another matter. As Deaton (2006) points out (in reviewing Fogel, 2004), human caloric requirements can be covered with seemingly modest spending on food staples.[113] However, this is not conclusive. Environmental enteropathy can generate quite low nutrition absorption rates given the persistent fecal–oral contamination of the environments in which many people live. In effect, the implicit price of an absorbed calorie capable of fueling work effort is higher, possibly far higher, than the nominal price. Furthermore, we have also learned that work productivity depends on the personal history of nutrition and health, as argued by Dasgupta (2011). Someone whose growth is stunted due to a long history of undernutrition—low intakes and/or low absorption—can be in current nutritional balance (able to afford current food energy requirements) but have such low productivity that a poverty trap emerges. It may not be a strict threshold, as in Figure 22.1, but a smoother, S-shaped function.

Other sources of threshold effects are also plausible on *a priori* grounds, such as the fact that a minimum level of schooling is essential before schooling can be a viable route out of poverty (recalling the story of Sunil in Boo, 2012). One can also interpret the aforementioned arguments about how poverty reduces cognitive functions as stemming from biological threshold effects—that a minimum level of time not worrying about the financial and other stresses created by poverty is needed to escape poverty (Mani et al., 2013).

In testing for threshold effects, some of the literature has looked for lumpiness in nonhuman capital requirements. The results have been mixed. Mesnard and Ravallion (2006) find evidence of nonlinear wealth effects on new business start-ups in Tunisia, but they do not find signs of thresholds effects. Nor do McKenzie and Woodruff (2006) find any sign of nonconvexities in production at low levels among Mexican microenterprises. In one of the few studies using wealth data, Barrett et al. (2006) do find evidence of the nonconvexity in asset data for rural Kenya and Madagascar.[114]

It can also be difficult to detect theoretically plausible threshold effects on dynamics in standard microdata sets (Day, 1992). For one thing, depending on the frequency of the observations over time in the data, the existence of the unstable "middle" equilibrium (point B in Figure 22.1) can generate attrition—the destitute simply drop out of the data (including by becoming homeless) (Lokshin and Ravallion, 2004). Also, there will be high social returns and risk-sharing arrangements to prevent most people falling into the trap. The trap is still there, but it may only be evident in extreme situations when those social relationships break down, as Ravallion (1997b) argues is the case during famines.

A testable implication of the models based on credit market failures is that individual wealth should be an increasing concave function of its own past value. In principle, this

[113] Subramanian and Deaton (1996) calculate that nutritional requirements can be met with a small fraction of the daily wage rate, using data for India. Similar reasoning leads Swamy (1997) to question the nutrition-based efficiency wage hypothesis.

[114] Also see the discussion in Carter and Barrett (2006).

can be tested on suitable micropanel data, though most data sets only show consumption or income, not wealth. Lokshin and Ravallion (2004) provide supporting evidence of concavity in panel data on incomes for Hungary and Russia while Jalan and Ravallion (2004) do so using panel data for China. These studies do not find the threshold properties in the empirical income dynamics that would be needed for a poverty trap. Using similar methods, but arguably a better identification strategy, Dercon and Outes (2013) find evidence of a low, unstable equilibrium in the income dynamics for a long panel of households in rural India.

Microempirical support for the claim that there are efficiency costs of poor nutrition and health care for children in poor families has come from a number of studies. In a recent example, an impact evaluation by Macours et al. (2008) of a conditional cash transfer (CCT) scheme in Nicaragua found that randomly assigned transfers to poor families improved the cognitive outcomes of children through higher intakes of nutrition-rich foods and better health care. This echoes a number of findings about the benefits to disadvantaged children of efforts to compensate for family poverty.[115]

The upshot of all this data is that present-day thinking is both more optimistic about the prospects of eliminating poverty through an expanding economy and more cognizant of the conditionalities in the gains to poor people from economic growth. Under the right conditions, growth can be a powerful force against poverty. Those conditions pertain in large part to aspects of both the initial distribution and how it evolves. As we will see in the following section, the focus of much antipoverty policy has shifted over time toward efforts to ensure that the conditions in place will allow poor people to contribute to an expanding overall economy, and so escape poverty permanently.

22.9. DIRECT INTERVENTIONS IN MODERN TIMES

If all incomes are observable and there are no behavioral responses, then guaranteeing a minimum income is straightforward—one simply makes transfers sufficient to bring everyone up to that minimum. Administrative capabilities, constraints on information, and incentive effects have meant that the practice of social policy is far more complicated. A range of interventions has emerged. This section discusses some generic issues—information, incentives, and policy design—before reviewing the main types of direct interventions found today.[116]

22.9.1 Generic Issues

The stage of development influences the types of policies needed. Poor places tend as a rule to have weaker administrative capabilities, which tends to mean less reliable

[115] For reviews of this literature, see Currie (2001, 2012).
[116] This section summarizes material from a much fuller discussion of antipoverty policy in Ravallion (2015).

information for deciding who should receive help. More universal (probably state-contingent) and/or self-targeted policies can thus have greater appeal in developing countries (including when the rich countries of today were developing), notably when there is a large informal sector. By contrast, the income tax system and means-tested transfer payments that require formalization tend to dominate in rich countries.

The existence of a large informal sector is associated with both information and incentive constraints on social policy in developing countries. The information constraints are obvious, given that informality essentially means that one has little systematic data about actual or potential beneficiaries. The incentive constraint comes from the fact that the informal sector is a feasible option for anyone in the formal sector (though the converse is less true). Thus, a social policy that can apply only to a formal-sector worker will have an added efficiency cost (through the scope for substitution) that would not be the case in a purely formal, developed economy.[117]

Incentive effects have figured in the debates about all forms of targeted direct interventions across all settings. A perfectly targeted set of transfers to poor families in the imaginary world of complete information—meaning that the transfers exactly fill the poverty gaps and so bring everyone up to the desired minimum income—would impose 100% marginal tax rates on recipients. This is very unlikely to be optimal from the point of view of poverty reduction given labor supply responses. One hundred and forty years after the famous debates over the reforms to England's Poor Laws, a rigorous formulation of the problem of redistributive policy with incentive effects was finally available in the form of Mirrless's (1971) optimal tax model. The Mirrless objective function was utilitarian, but his approach could also be adapted to an explicit poverty reduction objective. The simulations by Kanbur et al. (1994) suggested that marginal tax rates around 60–70% would be called for in an optimal antipoverty policy using transfers, allowing for incentive effects.[118]

At the opposite extreme to perfect targeting one can imagine a *basic income scheme*, which provides a fixed cash transfer to every person, whether poor or not.[119] This has been advocated by (among others) Paine (1797), Rhys-Williams (1943), Meade (1972), Raventós (2007), and Bardhan (2011). The idea has spanned both rich and poor countries, and the political spectrum from left to right. There are no substitution effects of the transfers because there is no action that anyone can take to change their transfer receipts, but there will be income effects (including higher demand for leisure, though how much so is unclear). There is no stigma associated with participation, given that there is no purposive targeting to poor people. A complete assessment of the implications

[117] Similarly, informal sector firms can evade taxation by resorting to cash (Gordon and Li, 2009).

[118] Also see Kanbur and Tuomala (2011) on alternative characterizations of the policy objective.

[119] This has been called many things including a "poll transfer," "guaranteed income," "citizenship income," and an "unmodified social dividend."

for efficiency (and equity) must take account of the methods of financing the scheme. The administrative cost would probably be low though certainly not zero given that some form of personal registration system would probably be needed to avoid "double dipping" and to ensure that larger households receive proportionately more money. Proposals in developed countries have typically allowed for financing through a progressive income tax (such as in Meade, 1972), in which case the idea becomes formally similar to the Negative Income Tax (Friedman, 1962) though the mode of administration may differ. Atkinson and Sutherland (1989) demonstrate that a basic income scheme can be devised as a feasible budget-neutral way of integrating social benefits and income taxation in Britain. In poor countries, a basic income scheme could be costly, depending on the benefit level and method of financing, although there may well be ample scope for financing by cutting current subsidies favoring the nonpoor, as Bardhan (2011) argues is the case for India. This type of scheme would appear to dominate many policies found in practice today; for example, it would clearly yield a better incidence than subsidies on the consumption of normal goods, which is a type of policy still found in a number of countries. However, as yet there have been very few examples of universal uniform cash transfer schemes in practice. (An example in Bolivia is discussed below.)

The bulk of the direct interventions found in practice fall somewhere between the above extremes of "perfect targeting" and a basic income with no targeting. In countries where income means testing is a feasible option (mostly rich countries), the benefit level can be progressively phased out as income rises above some level, below which some guaranteed support is provided. The rate of benefit withdrawal depends on the strength of the expected labor supply response. With the better data and analytic tools available today, it can be hoped that future policy debates will be better informed about actual behavioral responses. However, from what we know already about labor supply responses, it is evident that poor people gain significantly from transfers in a country such as the United States (Saez, 2006).

The recent emphasis on targeting in many countries (both rich and poor) has typically been defined as avoiding "leakage" of benefits to the nonpoor, implicitly downplaying concerns about coverage of the poor (as pointed out by Cornia and Stewart, 1995). Readily measurable proxies for poverty are widely used for such targeting in settings in which income means-testing of benefits is not an option. Efficiency considerations point to the need to use indicators that are not easily manipulated by actual or potential beneficiaries, although this is rarely very clear in practice. Geographic proxies have been common, as has gender of the recipient, family size, and housing conditions.[120] These targeting methods can be thought of as a "proxy means test" (PMT) in which transfers are allocated on the basis of a score for each household that can be interpreted as predicted

[120] Grosh et al. (2008) provide a useful overview of the targeting methods found in practice in developing countries, with details on many examples.

income or consumption, based on readily observed indicators. Depending on how it is designed, this type of scheme can have better incentive effects than perfect means testing and have a higher impact on poverty for a given outlay than a poll transfer. The main alternative method of targeting found in practice uses communities themselves to decide who is in greatest need. This exploits local information that is not normally available for the PMT, but it does so at the risk of exploitation by local elites.[121] However, policy advisors and policy makers sometimes appear to have treated "better targeting" as the objective of the policy design problem, forgetting that it is really only an instrument, and not necessarily the best instrument given the aforementioned costs and the political economy response to targeting, whereby finely targeted programs can undermine the political support for social policies.[122]

22.9.2 State-Contingent Transfers Financed by Taxation

Recall that the essential idea of England's Old Poor Laws was state-contingent transfers financed by taxation. There was little effort at explicit targeting of relief (before the 1834 reforms, which we return to later), although there was some degree of self-targeting given that relatively well-off families would be reticent to turn to the parish for assistance after some economic shock.

The idea of untargeted state-contingent transfers (as in the Old Poor Laws) reemerged in twentieth century Britain in the form of the *Beveridge Report* (Beveridge, 1942), which outlined detailed proposals for social insurance, whereby all those of working age would be obliged to pay a national insurance contribution to finance state-contingent transfers to the unemployed, the sick, the elderly, or widowed. However, unlike the Old Poor Laws, this was to be a national scheme rather than implemented locally. Two other elements completed the social protection policy. First, family allowances were proposed, to cover the costs of dependent children (after the first). Second, an income top-up was proposed for those who fell below absolute standards taking account of all income sources.[123] Although the aim of these proposals was squarely to eliminate poverty, Beveridge was opposed to means-testing—universal provision at a flat-rate was seen to avoid the costs of targeting and to encourage social cohesion.[124] The past, deliberately stigmatizing,

[121] Discussions of community-based targeting can be found in Alderman (2002), Galasso and Ravallion (2005), Mansuri and Rao (2012), and Alatas et al. (2012). The latter paper compares this form of targeting with PMT for a cash transfer program in Indonesia. The study finds that PMT does somewhat better at reaching the poor, but community-based targeting better accords with local perceptions of poverty and is better accepted by local residents.

[122] For further discussion, see van de Walle (1998), De Donder and Hindriks (1998), and Gelbach and Pritchett (2000).

[123] This came to be known as the "Supplementary Benefit" and became more important in practice than Beveridge envisaged; see the discussion in Meade (1972).

[124] There is an interesting discussion of Beveridge's arguments in Thane (2000, especially, Chapter 19).

approach typified by the workhouses was to be abandoned. Beveridge's plan formed the basis for the policies of the new Labour government elected in 1945; the Conservative resistance to the (popular) Beveridge plan helped ensure a Labour victory (Thane, 2000, p. 369).

America's Social Security system had also grown out of prior relief efforts (notably those established during the depression) and came to provide fairly comprehensive state-contingent transfers, financed by taxation, soon after World War II. As with the Poor Laws, there was much debate about these policies. (America's Social Security system was, and still is, decried as "socialism" in some quarters.) Similar to the 1834 reforms to the Poor Laws, calls for targeting have become common since 1980 in an attempt to reduce the fiscal cost of social insurance.

Uniform but state-contingent transfers are not common in developing countries today. It seems that developing countries have largely skipped this stage in the history of social policy. However, it is not entirely clear why this is the case or that it is a good idea from the point of view of sound policy making. To explain why uniform state-contingent transfers of the social insurance type are not used, it is sometimes claimed that such policies are unsuitable to poor economies; they would be too costly, and targeting is needed. Although the fiscal burden of social policies must never be ignored, it is notable that the Old Poor Laws were invented in what was clearly a poor economy by today's standards. For some 300 years, the Old Poor Laws provided a degree of social protection and stability at a seemingly modest cost (Solar, 1995).

As we will see, although better targeting may help, finely targeted policies have costs that are often hidden but must be considered in any proper evaluation of the policy options.

22.9.3 Workfare

The workhouses that emerged in Europe around 1600 can be interpreted as a means of getting around the information and incentive problems of targeting. Design features encouraged those truly in need of help to turn to the workhouse *and* encouraged them to leave it when help was no longer needed, given that there were better options in the rest of the economy. This solves the information problem of targeting. However, it does so by imposing costs on participants, notably the forgone earnings and the welfare costs of stigma and subjugation (as Oliver Twist experienced). A truly utilitarian–welfarist assessment relative to untargeted transfers would clearly be ambiguous without further evidence. Arguably England's workhouses of the nineteenth century went too far in imposing costs, which came to be widely seen as objectionable, on participants to ensure self-targeting. But the idea of self-targeting had lasting influence.

The workhouses are an example of a class of direct interventions often called today "workfare schemes"—schemes that impose work requirements on welfare recipients as a means of ensuring incentive compatibility. Though not involving workhouses, this idea was embodied in the Famine Codes introduced in British India around 1880, and the idea has continued to play an important role to this day on the subcontinent (Drèze, 1990a). Such schemes have helped in responding to, and preventing, famines, including those in Sub-Saharan Africa (Drèze, 1990b). Workfare was also a key element of the New Deal introduced by US President Roosevelt in 1933 in response to the Great Depression.

An important subclass of workfare schemes has aimed to guarantee employment to anyone who wants it at a predetermined (typically low) wage rate. Employment Guarantee Schemes (EGSs) have been popular in South Asia, notably (though not only) in India where the Maharashtra EGS, which started in 1973, was long considered a model. In 2005, the central government implemented a national version, the Mahatma Gandhi National Rural EGS. This promises 100 days of work per year per rural household to those willing to do unskilled manual labor at the statutory minimum wage listed for the program. The work requirement is (more or less explicitly) seen as a means of ensuring that the program reaches India's rural poor.[125]

These schemes can be interpreted as attempts to enforce a minimum wage rate in situations in which there is no other means of legal enforcement. Minimum wages appeared in the late nineteenth century, with the first minimum wage law introduced by New Zealand in 1894. Critics have long pointed to concerns about negative effects on overall employment of minimum wages rates, although advocates have pointed out that those effects may be small in practice and even positive in monopsonistic labor markets. However, enforcement of minimum wage legislation has been famously weak in developing countries with large informal sectors (including traditional farming). For example, Murgai and Ravallion (2005) show that in 2004–2005, three-quarters of India's casual labor was paid less than the country's (state-level) statutory minimum wage rates. In an EGS, anyone who wants work can (in theory) get it provided they are willing to do unskilled manual labor at the statutory minimum wage rate in agriculture.

An important difference between an EGS and minimum wage legislation is that an EGS aims to provide comprehensive insurance for the able-bodied poor, in that anyone who needs work can get it, at least on paper. Eligibility is open to all, so that a farmer who would not need the scheme in normal times can turn to it during a drought (say). This concept was explicit from the outset of the idea of an EGS (as it developed in Maharashtra in the early 1970s). Whether this insurance function is served in practice is another matter; Dutta et al. (2012) find evidence of considerable rationing on India's national EGS. The rationing tends to be greater in poorer states, which may well reflect their weaker administrative capabilities to implement a complex program such as India's national EGS.

[125] Dutta et al. (2013) provide an assessment. Also see Jha et al. (2012), Gaiha (1997), and Imbert and Papp (2011).

These schemes illustrate that even a well-targeted transfer scheme can be dominated by untargeted transfers when one takes account of all the costs involved, such as income forgone or other costs in complying with the conditionalities imposed. Ravallion and Datt (1995) and Murgai et al. (2013) provide evidence that in both the Maharashtra EGS and India's new national scheme, an untargeted basic income scheme would have been more cost-effective in directly transferring incomes to poor people.

Workfare schemes have typically been seen as short-term palliatives—a form of social insurance. In principle, a workfare scheme can also directly serve promotional goals. One way is by generating assets that could change the wealth distribution or shift the production function, which could also allow people to break out of the poverty trap illustrated in Figure 22.1. In practice, asset creation has not been given much weight in these schemes in South Asia, although it seems to have greater weight elsewhere, including in Latin America (such as Argentina's Trabajar Program).

Another way that workfare programs can better serve the promotional goal is by tying benefits to efforts to enhance human capital through training. Welfare reforms in many rich countries since the early 1990s have aimed to make transfers conditional on investments in human capital and to incentivize searching for and finding private employment.[126] This form of workfare does not actually provide employment, as in the public works form of workfare. Training and encouragements for private-sector employment using wage subsidies have also been used to encourage the transition from public employment in workfare schemes to private employment.[127]

Next we turn to a policy for which the creation of human wealth is seen as crucial to poverty reduction.

22.9.4 Schooling for Children from Poor Families

Children from poor families tend to get less schooling. This "economic gradient" in schooling persists to this day almost everywhere and has long been seen as a factor that perpetuates poverty across generations—a potential source of a poverty trap. As noted in Section 22.5, the inability of poor families to finance their children's schooling given credit market failures came to be recognized as a key factor in perpetuating poverty and entailing that a more unequal initial wealth distribution will generate aggregate efficiency costs.[128] Thus, policies that can promote the schooling of children from poor families can be seen as an important part of social policy that could improve both equity and efficiency, and credibly allow people to escape poverty permanently.

[126] Hemerijck (2014) provides an overview of such reforms in Europe.

[127] An example is the Proempleo scheme in Argentina studied by Galasso et al. (2004).

[128] Physical accessibility is sometimes identified as another factor. However, the simulations by Filmer (2007) do not suggest that this is a major factor in the schooling gap between rich and poor in developing countries; Filmer finds that very large reductions in distance to school would be needed to close the gap.

Such policies are a modern idea, advocated at times but little known in practice before the nineteenth century (see Section 22.5). Past and ongoing policy debates over mass education have raised many issues, but a fundamental one is whether compulsory schooling is even in the interest of poor families, for it was typically their children who were unschooled. Opponents (on both the left and right) of compulsory schooling pointed to the costs (primarily their forgone earnings) to poor families of sending their children to school. While compulsory schooling could break the poverty trap, a short-term trade-off was created by the costs to poor families. Advocates argued, in effect, that the longer-term benefits from breaking out of a poverty trap outweighed these costs.

After much debate, compulsory schooling emerged in virtually all industrialized countries by the early twentieth century, with a significant state role in both public provision and support for private schooling. In England, the Elementary Education Act of 1870 was a breakthrough law that established a secular public sector institutional framework, including democratic school boards. Implementation was uneven geographically, and there was a continuing struggle for control of schools between the democratically elected local bodies and religious organizations (Stephens, 1998). It was not until the 1880 act of the same name that education was compulsory in England for children aged 5–10. A similar act was passed in France about the same time. In the United States, 34 states had compulsory schooling laws by 1900, 30 of which required attendance until at least age 14. Japan in the Meiji period (1868–1912) was not behind the West in promoting mass education, which was virtually universal by the end of the period. Mass public education (with tertiary education left largely to the private sector) was given high priority throughout developing East Asia, with educational attainments far surpassing those of most developing countries and even some developed countries.

The payoffs from mass public education were huge. Equitable, broad-based education has been identified by Goldin and Katz (2008) as a key factor in the US record of relatively equitable and rapid economic growth in the period 1940–1980. The ability of the school system to support a relatively rapid increase in education attainments in the United States in this period (though slowing down greatly after 1980) meant that the supply of skilled workers kept up with the extra demand stemming from new technologies—what Tinbergen (1975) dubbed the "race between education and technology"—thus attenuating the inequality-increasing effects of technical progress favoring demand for relatively skilled labor. The fact that American educational expansion was so broad-based in this period was key. A more elitist school system would have entailed a more unequal distribution of the gains from growth. And Goldin and Katz argue that rising inequality in the United States since 1980 stems in large part from the fact that the education system has not allowed the supply of the types of skilled labor required for the new technologies of the time to keep up with the demand. And it tends to be children from poor families who are most disadvantaged in this race.

Broad-based education has also been identified as a key factor in East Asia's relatively equitable growth. Using a regression of GDP per capita growth rates from 1960 to 1985 on primary and secondary education attainments in 1960—with controls for initial GDP per capita, population growth, and the share of investment in GDP—an influential report by the World Bank (1993) identified primary education as the most important single factor, accounting for somewhere between 58% (Japan) and 87% (Thailand) of GDP growth. Of course, such calculations can be sensitive to model specification; the education variables could well be correlated with other omitted factors. However, it is nonetheless striking that primary education is found to account for a greater share of the variance in growth rates than private (nonhuman) investment.

There is also evidence that education attainments have interacted strongly with India's growth process in determining the impact of that growth on poverty. This was demonstrated by Ravallion and Datt (2002) by comparing rates of poverty reduction among India's states. While the elasticities of measured poverty to farm yields did not vary significantly across states, those for nonfarm output did. The nonfarm growth process tended to reduce poverty more significantly in states with initially higher literacy rates, and interstate differences in literacy rates were the dominant factor among those identified by Ravallion and Datt. The importance of mass education has long been acknowledged in principle in India. A "directive principle" of state policy in the 1949 Constitution was free compulsory education to the age of 14.[129] However, implementation of this policy has lagged considerably, with large interstate differences and often poor quality schooling across the country (Probe Team, 1999). The state that has made the most progress in mass public education is Kerala. Expanding literacy to the whole population was a high priority of the state government from the 1950s (building on a history of prior successes in schooling provided by Christian missionaries dating back to the early nineteenth century). The results of Ravallion and Datt (2002) indicate that Kerala's success in mass schooling has generated a far more propoor process of nonfarm economic growth than is found in other states.

Bans on child labor have often been proposed and legislated. Hazan and Berdugo (2002) model an interesting version of a poverty trap in which, at the early stage of development, child labor is abundant while fertility is high, and mean output is low. With economic growth stemming from technical progress, the returns to schooling rise, making child labor less attractive and also lowering fertility. In this model, the economy eventually converges to a new equilibrium in which child labor has vanished. Hazan and Berdugo show that an effective ban on child labor will speed up the transition to this new equilibrium.

However, in economies with large informal sectors, the enforcement of such bans is difficult. Legislation to set a minimum working age was introduced in some countries

[129] A Right to Education Act was passed by India's parliament in 2009, essentially ratifying the Constitution.

from the late nineteenth century, although it is unclear how much this helped reduce the incidence of child labor; Moehling's (1999) analysis suggests this legislation had little effect. Basu (1999) argues that compulsory schooling is a better way of implementing a ban on child labor than an actual ban, and compulsory schooling can also break the poverty trap.

22.9.5 Policy Incentives for Schooling

Although out-of-pocket expenses and the forgone earnings of children figured in the nineteenth-century debates about the idea of compulsory education, there was not much discussion of the obvious policy response: a bursary, or scholarship grant, for poor families. Smith (1776) and Mill (1859, Chapter 5) had advocated tuition subsidies for children from poor families. Marshall (1890, p. 594) took a less sympathetic attitude and proposed instead penalizing poor parents (a public policy of "paternal discipline") who neglected to send their children to school or to care for their health. Educational institutions have for a long time subsidized tuition fees and other costs for selected students, often based on some sort of means test. England's 1870 Elementary Education Act recommended tuition subsidies for children from poor families (Gillie, 1996). However, implementation of public policies providing any form of schooling incentive for poor parents had to wait until the middle of the twentieth century, after which it started to become common practice to build in incentives for children from poor families to stay in school. Britain's 1942 *Beveridge Report* recommended a universal child allowance paid up to the age of 16 if the child stayed in school.[130] Australia had a school bursary program from the 1960s that essentially paid parents from poor families to keep their children in school beyond the age the children would normally leave school as long as the children passed a special exam. It is common today for various forms of education subsidies (scholarships, tuition subsidies, subsidized loans) to be means-tested.

In the development literature in the 1990s, targeted bursaries came to be known as CCTs.[131] The idea was the same: a monetary incentive for poor parents to keep their children in school. Transfers are made under the condition that the children of the recipient family demonstrate adequate school attendance (and health care in some versions of the policy). Plainly, the promotion benefits of these programs rest on ensuring that the transfers go to poor families, presuming that the children of the nonpoor will already be in school. Thus, targeting has been instrumentally important to both the protection and promotion benefits. The promotion benefits also depend on designing the conditions so that the required level of schooling would not be attained in the absence of the program. Early influential examples of these programs in developing countries were

[130] Similarly, the US Earned Income Tax Credit (introduced in 1975) gives different age cut offs for full-time students.

[131] Most other direct interventions also have conditions; for example, workfare entails a work requirement.

Mexico's PROGRESA program (now called Oportunidades) and Bolsa Escola in Brazil. Another early example was the Food-for-Education Program in Bangladesh for which the transfers (targeted to poor families) were made in kind but were also conditional on school attendance. Bolivia's CCT, Bono Juancito Pinto, introduced in 2006, is an example of a universal (untargeted) transfer program, for which every child enrolled in public school is eligible, irrespective of family income. More than 30 developing countries now have CCT programs, and the number is growing (World Bank, 2014). And other countries have similar policies that are not called CCTs; for example, in an attempt to ensure that poverty did not constrain schooling, since 2002 China has had a "two exemptions, one subsidy" policy for students from poor rural families; the exemptions are for tuition fees and textbooks, and the subsidy is for living costs.

These programs are clearly designed with a view to breaking the poverty trap stemming from the aforementioned economic gradient in human development. If the sole concern was with current income gains to participating households, then a policy maker would not impose schooling requirements, which entail a cost to poor families by incentivizing them to withdraw children or teenagers from the labor force, thus reducing the (net) income gain to poor people. The idea of these programs is to strike a balance between protection and promotion, based on the presumption that poor families cannot strike the socially optimal balance on their own. The program's incentive effect on labor supply (previously seen as an adverse outcome of transfers) is now judged to be a benefit—to the extent that a well-targeted transfer allows poor families to keep the kids in school, rather than sending them to work. Concerns about distribution *within* households underlie the motivation for such programs; the program's conditions entail that relatively more of the gains accrue to children. Some advocates of CCTs have also claimed that they would reduce child labor, although the economic data are unclear about whether such a policy will work for this purpose; Ravallion and Wodon (2000a) show that, under standard assumptions, a tuition subsidy will increase children's amount of schooling but has theoretically ambiguous effects on the supply of child labor; empirically, the authors find that a tuition subsidy has little effect on child labor in Bangladesh.

There is evidence from impact evaluations that these schemes bring nonnegligible benefits to poor households, in terms of both current incomes and future incomes, through higher investments in child schooling and health care.[132] The conditions change behavior. In the United Kingdom, means-tested grants paid to secondary students have been found to very effectively reduce the number of school drop outs from poor families (Dearden et al., 2009). The various evaluations of Mexico's PROGRESA/Oportunidades program have been positive; see the survey in Fiszbein and Schady (2010). Baird et al. (2011) found sizeable gains from the schooling conditions in a Malawi CCT. In a

[132] Fiszbein and Schady (2010) provide a comprehensive review. Also see the discussion in Das et al. (2005).

study for Burkina Faso, Akresh et al. (2013) found that the conditionality mattered more in encouraging the school enrollment of children who were initially less likely to go to school, including girls—children who are less likely to receive investments from their parents. Cameron (2002) found that a CCT program in Indonesia, Jaring Pengamanan Sosial, had the greatest impact at the lower secondary school level where children are most susceptible to dropping out. The design features have also been critically assessed. A series of papers on PROGRESA revealed that a budget-neutral switch of the enrollment subsidy from primary to secondary school would have delivered a net gain in school attainments by increasing the proportion of children who continue onto secondary school.[133] Although PROGRESA had an impact on schooling, it could have had a larger impact. However, it should be recalled that this type of program has two objectives: promotion by increasing schooling (reducing future poverty) and protection by reducing current poverty through the targeted transfers. To the extent that refocusing the subsidies on secondary schooling would reduce the impact on current income poverty (by increasing the forgone income from children's employment), the case for this change in the program's design would need further analysis.

Impact evaluations have also pointed to high returns for early childhood interventions in some settings. The experimental Perry Preschool Program in the United States in the 1960s provided schooling and home visits for poor children aged 3–4. The benefits included higher adult earnings and reduced crime, and the benefit-cost ratio (even without putting greater weight on the propoor distribution of the gains) was estimated to be more than eight to one (Heckman, 2006). Head Start (also begun in the 1960s as part of the United States's War on Poverty) was a similar national preschool program, which targeted a package of education, health, and nutrition services to poor families; the program continues at the time of this writing and, as of 2005, some 22 million preschool children had participated in Head Start programs. Head Start has also been found to generate sizeable long-term gains in schooling, earnings, and reduced crime (Garces et al., 2002). The aggregate benefits from Head Start also appear likely to exceed the cost, even without distributional weights (Ludwig and Phillips, 2007). There is also evidence of significant long-term gains in adult health indicators from an intensive preschool program launched in the United States during the 1970s, the Carolina Abecedarian Project (Campbell et al., 2014). There is a great deal of interest in how effective early childhood interventions might be devised for developing countries.

All these interventions require complementary efforts on the supply side, through effective (public or private) service delivery. This has been an important concern in many developing countries; World Bank (2004) reviews the evidence and discusses how better incentives for service delivery might be developed.

[133] See Todd and Wolpin (2002), Attanasio et al. (2004), and de Janvry and Sadoulet (2006).

22.9.6 Microfinance Schemes

As we have seen, credit market failures have been identified as a cause of poverty and a reason why poverty can be costly to overall economic performance. In addition to long-standing moral arguments, transfers to poor people can be interpreted as a means of relieving the constraints stemming from such market failures. But there is another option, namely, policies that aim to make financial institutions for saving and borrowing work better for poor people. Microfinance programs aspiring to support small-scale credit and savings transactions by poor people have attracted a great deal of interest since the idea emerged in the late 1970s, and there are now many examples of such programs in the developing world.

The classic argument is about promotion, namely, that relaxing borrowing constraints faced by poor people allows them to invest and so eventually escape poverty by their own means. Credit and savings are also potentially important instruments for protection, by allowing poor households to more effectively smooth their consumption in the face of income fluctuations.

Much of the early (and ongoing) enthusiasm for microfinance was really little more than advocacy, with weak conceptual and empirical foundations. In recent times, there has been a rise in popular concern in the media (in South Asia especially) about over-borrowing by poor people once they are given access to microfinance. Much of this concern appears to stem from anecdotes, and the debate has also become politicized. Positive average impacts do not, of course, mean that there are no losers among the recipients. This is probably true of all antipoverty policies, but it is especially so in the case of credit-based interventions. Risk is not eliminated, shocks do occur, and mistakes are made, such as due to faulty expectations. There will be both winners and losers in these types of interventions.

The earliest and still most famous example of this class of policies is Bangladesh's group-based lending scheme, Grameen Bank (GB). GB has made a conscious effort to reach the poor both through their eligibility criteria and their branch location decisions, which (in contrast to traditional banks) have favored areas where there are unexploited opportunities for poor people to switch to nonfarm activities (Ravallion and Wodon, 2000b). Research on GB has indicated that the scheme has helped in both protection and promotion; in the former case, by facilitating consumption smoothing and, in the latter, by helping to build the physical and human assets of poor people.[134] This result was found by Pitt and Khandker (1998), who exploited the design features of GB, notably that it is targeted to the landless, to identify effects. Given that access to GB raises the returns of being landless, the returns of having land will be higher in villages that do not have access to GB credit. Thus, comparing the returns of having land between

[134] An early contribution to knowledge about GB was made by Hossain (1988).

villages that are eligible for GB and those that are not (with controls for other observable differences) reveal the impact of access to GB credit. Put another way, Pitt and Khandker measure the effect by the mean gain among households that are landless from living in a village that is eligible for GB, less the corresponding gain among those that have land. They found positive impacts on measures relevant to both protection and promotion. This was confirmed in a subsequent study by Khandker and Samad (2014) using survey data on 3000 households spanning 20 years. The success of GB has led to a proliferation of microfinance schemes in Bangladesh, with over 500 providers at the time of this writing.

Even careful observational studies such as that by Pitt and Khandker require identifying assumptions that can be questioned, and there has been some debate about the robustness of their results.[135] This is a type of policy intervention for which it will inevitably be hard to convince everyone of the validity of the identifying assumptions given the likelihood of unobservable factors jointly influencing acceptance and effects. Experimental evaluations relying on randomized assignment (typically at the community level) have offered the hope of more robust results, and there have been some interesting examples. A study by Banerjee et al. (2009) of the impacts of opening new microfinance branches in slums of Hyderabad in India found that overall borrowing, business start-ups, and spending on consumer durables (but not nondurables) increased in the areas that were randomly assigned the new branches relative to the control areas. However, the study did not find evidence of positive impacts on health, education, or women's self-efficacy. Heterogeneity was the focus of a recent experimental evaluation of access to micro-credit by working-age women in Mexico (under the Compartamos Banco scheme) by Angelucci et al. (2013). The authors found positive average effects in a number of dimensions. There was heterogeneity in the impacts, but they found little evidence of significant losses, including among poor borrowers. More research on the benefits and costs of microfinance schemes can be expected.

It is clear that we have seen a shift in thinking about this class of policies over the last 200 years; in the days when poor people were routinely blamed for their poverty, giving them a loan would not have made sense. Of course, identifying credit market failures as one cause of poverty does not imply that credit for the poor will solve the problem. But well-designed programs do have a role, as a complement to other policies for protection and promotion.

22.9.7 Poor-Area Development Programs

Almost all countries have their well-recognized "poor areas," in which the incidence of absolute poverty is unusually high by national standards. We would hope, and under certain conditions expect, that the growth process will help these poor areas catch up.

[135] See Morduch (1999) and Roodman and Morduch (2009) as well as the latest detailed rejoinder in Pitt and Khandker (2012).

But this process often appears to be slow, and geographic divergence has sometimes been evident. This has led to antipoverty policies focused on lagging poor areas. "Poverty maps" are widely used in geographic targeting, and the method proposed by Elbers et al. (2003) has facilitated many applications.

Lagging poor areas have prompted poor-area development projects—one of the oldest forms of development assistance, though under various headings (including "Integrated Rural Development Projects" and "Community Driven Development"). Extra resources are channeled to the targeted poor areas for infrastructure and services and developing (farm and nonfarm) enterprises. Emphasis is often given to local citizen participation in decision making, although a survey of the available evaluative research by Mansuri and Rao (2012) found somewhat mixed success given the scope for exploitation by local elites.

It is widely agreed that poor areas are typically characterized by low capital-to-labor ratios, but there is less agreement about the right policy responses, such as efforts to augment local capital—investing in lagging poor areas—versus policies to encourage out migration. Geographic externalities clearly play an important role, but this role is still poorly understood because of a lack of convincing empirical research.

In the case of China, where poor area development has been the main form of direct intervention against poverty since the mid-1980s, there is evidence of pervasive geographic externalities, whereby households living in poor areas have lower growth prospects than seemingly identical households living in well-off areas (Jalan and Ravallion, 2002; Ravallion, 2005). This suggests that there is scope for poor-area development as a means of ensuring longer-term promotion from poverty, as well as protection. However, here, too, the evidence for the success of the policies currently in practice is mixed.[136]

The main concerns about the incentive effects of poor-area programs have related to the responses of local governments to external aid and to migration. An example of the former is found in Chen et al. (2009), who demonstrate that local government spending allocations changed in response to efforts by higher levels of government to target poor villages in rural China, dampening the targeting outcomes. Regarding migration, there appears to be a widely accepted assumption that there is limited intrarural mobility in developing countries, sometimes reflecting institutional and policy impediments (such as local administrative powers for land reallocation, as in China). It is not clear how confident we can be in making that assumption.

There is still much we do not know about the impacts of poor-area development efforts, especially over the long term, and the tradeoffs faced by policy options. Although local infrastructure development is clearly crucial to fighting poverty, it has not attracted the degree of attention in evaluative research that has been generated by social policies. Here, an important factor is the extent to which "development impact" is challenged by

[136] For example, contrast the findings of Jalan and Ravallion (1998) with Chen et al. (2009) on poor-area programs in China.

donors and citizens. "Impact" is too often taken for granted with infrastructure development. By contrast, the "softer" social policies have had to work hard to justify themselves, and evaluative research has served an important role. If the presumption of impact is routinely challenged by donors, aid organizations, and citizens, then we will see stronger incentives for learning about impact and fewer knowledge gaps.

22.9.8 Information Campaigns

There has been recent interest in the scope for using information-based interventions because lack of information is a decisive factor inhibiting poor people from successfully participating in actions to get the services to which they are entitled. There are some signs of support for this premise from past research. Strömberg (2004) reports evidence that US antipoverty programs have worked better in places with greater access to radios. Besley and Burgess (2003) found that the governments of Indian states where newspaper circulation is greater are more responsive in their efforts to mitigate negative agricultural shocks. Reinikka and Svensson (2005) found significant effects of information through a newspaper campaign on school outcomes in Uganda.

There have been some evaluations of information interventions. The results so far seem mixed. Focusing on one country and one sector, Pandey et al. (2009) report that a community-based information campaign led to short-term gains in schooling outcomes, while the findings of Banerjee et al. (2010) are less encouraging about the scope for using information interventions to improve the monitoring of education service providers in India. In rich countries facing concerns about the rising incidence of obesity, there have been efforts to post information on the "calorie prices" of food.[137] A recent review of both experimental and nonexperimental evaluations found mixed evidence of effectiveness (Swartz et al., 2011).

Mixed results of this sort might not be surprising. Three observations can be made. First, public information about a program may well discourage participation; for some people, learning about the program may have the opposite effect; see, for example, Hertel-Fernandez and Wenger (2013), with regard to an information campaign for a US program. Second, incomplete information is only one of the possible reasons why poor people do not access services (Keefer and Khemani, 2005; Cappelen et al., 2010). Third, mixed results might also stem from heterogeneity in the quality of the information intervention itself. Also, for India, Ravallion et al. (2013) report success in changing public awareness of rights and rules under India's EGS using an entertaining and high-quality fictional movie that can be shown in villages. However, the results also warn that informing poor people of their rights is not sufficient for positive change. Public awareness can be improved, but this must be combined with effective responses on the supply side.

[137] For example, US legislation in 2010 requires restaurant chains with 20 or more outlets to post calorie counts for all food items sold.

22.10. CONCLUSIONS

This chapter has tried to describe and better understand how the idea of antipoverty policy emerged and has evolved over the last 200 years. It has been argued that we have transitioned from one view of poverty to another, radically different view of poverty. In the first, there was little reason to think that poor people had the potential to be anything else than poor. Poverty would inevitably persist, and, indeed, it was deemed necessary for economic expansion, which required a large number of people eager for work; avoiding hunger was seen as the necessary incentive for poor people to do that work. Social policy had a role in ensuring social stability—most importantly, a generally docile working class willing to work for low wages—and successfully so it seems in the case of England's Poor Laws. Promotional antipoverty policies would probably not have made much sense to those in power, although the need for protection from economic shocks would have been more evident and appears to have had reasonably broad support from the elites even when mass chronic poverty was taken for granted. However, beyond short-term palliatives to address shocks, there was little or no perceived scope for public efforts to permanently reduce poverty. And a world free of poverty was unimaginable—after all, who then would be available to farm the land, work the factories, and staff the armies?

In the second, modern, view, poverty is not only seen as a social ill that can be avoided through public action, but such action is seen as perfectly consistent with a robust growing economy. Indeed, the right antipoverty policies are expected to contribute to development by removing material constraints on the freedom of individuals to pursue their own interests.

Granted, such a public commitment is not universal today in any country. Some observers still point to behaviors of poor people as causes of their poverty, while others point to constraints beyond their control. Advocates against poverty are often frustrated by the setbacks. However, the progress that has been made in both the idea of antipoverty policy and its effective implementation is undeniable. Recognizing such a marked transition in mainstream thinking over 200 years makes one more optimistic that the idea of eliminating poverty can be more than a dream.

Progress has been uneven over time. Two key historical steps in the transition can be identified, dubbed here as the First and Second Poverty Enlightenments. The First, taking place just before the turn of the nineteenth century, saw the emergence of a new respect for poor people as people—no longer the "shadows in a painting" or objects that served some purely instrumental role as means of production. Instead, the economy itself came to be seen as a means to promote human welfare, including that of poor people. The Second Poverty Enlightenment, in the latter part of the twentieth century, came with the strongest case yet for antipoverty policy, which saw poverty as a severe constraint on freedom and personal self-fulfillment. A consensus emerged that poverty was morally unacceptable, though with continuing debates on what to do about it.

Although the foundation for this change was laid in the First Poverty Enlightenment—notably in seeing all human beings as morally equal, with legitimate desires for freedom and self-fulfillment—it was really only by the time of the Second Poverty Enlightenment that it came to be understood that freedom and self-fulfillment required (among other things) that people were not constrained by poverty. The state was seen to have a role in ensuring that all individuals had access to the material conditions for their own personal fulfillment—arguably the most important requirement for equity, but also the key to breaking out of poverty traps. Antipoverty policy came to be seen as a matter of both promotion and protection. Along with rising real wages and (hence) savings by poor people, public education systems, sound health care systems, and reasonably well-functioning financial markets came to be seen as crucial elements for the next generation of poor families to escape poverty for good.

Once it started to be widely accepted that those born poor could in fact escape poverty, public action against poverty became more acceptable, and more people joined political coalitions or struggles toward that end. Once successful promotion policies had been initiated, the fiscal burden of providing relief to those who remained poor started to fall. This was probably reinforced by new political support for action and moral conviction about its need, stemming from the world's (now much expanded) middle class. Beyond some point, a self-reinforcing cycle emerged in the successful countries to help ensure a sustained and (over time) more rapid escape from poverty. The cycle has been broken at times; the history of thinking and action on poverty gives ample illustration of the fragility of the progress we have seen. Each Poverty Enlightenment was followed by a backlash in thinking and policy making. But we have seen progress.

ACKNOWLEDGMENTS

This chapter's title owes a debt to Gertrude Himmelfarb's (1984) text, *The Idea of Poverty: England in the Early Industrial Age*. For comments the author thanks Robert Allen, Tony Atkinson, Pranab Bardhan, Francois Bourguignon, Denis Cogneau, Jean-Yves Duclos, Sam Fleischacker, Pedro Gete, Karla Hoff, Ravi Kanbur, Charles Kenny, Sylvie Lambert, Philipp Lepenies, Peter Lindert, Michael Lipton, Will Martin, Alice Mesnard, Branko Milanovic, Johan Mistiaen, Berk Ozler, Thomas Pogge, Gilles Postel-Vinay, Henry Richardson, John Roemer, John Rust, Agnar Sandmo, Amartya Sen, Dominique van de Walle, and participants at presentations at the World Bank, the Midwest International Economic Development Conference at the University of Wisconsin–Madison, the Paris School of Economics, the Canadian Economics Association, the 12th Nordic Conference on Development Economics, Lancaster University, and the World Institute for Development Economics Research.

APPENDIX

This Appendix proves the claim made in Section 22.8 about the properties of the characterization of wealth dynamics in Section 22.2. The claim in Section 22.8 referred to a situation in which the threshold is not binding, giving the Banerjee and Duflo (2003) model. The latter paper shows that higher initial wealth inequality lowers future growth in wealth. Here, we focus instead on the implications of high initial wealth poverty.

Initial wealth, w_t for date t, is distributed across individuals according to the cumulative distribution function, $F_t(w)$, giving the population proportion with wealth lower than w, and let $H_t = F_t(z)$ denote the headcount index of poverty (poverty rate) when the poverty line is z. (It will be analytically easier to work with the inverse of $F_t(w)$, namely, the quantile function, $w_t(p)$.) If credit is constrained ($w_t \leq k^\star/(\lambda+1)$), then output at $t+1$ is limited by the amount of capital available at time t, which is given by own-wealth plus maximum borrowing, yielding an output of $h((\lambda+1)w_t)$. The recursion diagram for the credit-constrained individual then takes the form:

$$w_{t+1} = \varphi(w_t) = \beta[h((\lambda+1)w_t) - r\lambda w_t] \quad \text{for} \quad w_t \leq k^*/(\lambda+1) \tag{23.1}$$

By contrast, the following recursion diagram holds for the unconstrained person (who is free to implement the optimal capital stock $k^\star$ at which point $h'(k^\star) = r$):

$$w_{t+1} = \beta[h(k^*) + r(w_t - k^*)] \quad \text{for} \quad w_t > k^*/(\lambda+1) \tag{23.2}$$

Here, β is the fixed share of current wealth that is not currently consumed. Plainly, $\varphi(w_t)$ is strictly concave up to $k^\star/(\lambda+1)$ and linear above that point. It is assumed that $z \leq k^\star/(\lambda+1)$. Let $H_t^* \equiv F_t[k^\star/(\lambda+1)]$. Mean future wealth is:

$$\mu_{t+1} = \int_0^\infty \varphi[w_t(p)]\mathrm{d}p \tag{23.3}$$

By standard properties of concave functions, we can readily verify that an inequality-increasing spread in the wealth distribution in this economy will reduce mean future wealth at a given level of mean current wealth, that is, reduce the growth rate, as in Banerjee and Duflo (2003).

What about the impact on growth of higher initial poverty at a given initial mean? Using Equations (23.1) and (23.2), we can rewrite Equation (23.3) as:

$$\mu_{t+1} = \beta\int_0^{H_t^*} [h((\lambda+1)w_t(p)) - \lambda r w_t(p)]\mathrm{d}p + \beta\int_{H_t^*}^1 [h(k^*) + (w_t(p) - k^*)r]\mathrm{d}p \tag{23.4}$$

Consider the growth effect of a mean-preserving increase in the poverty rate. It is assumed that H_t^* increases and that no individual with wealth less than $k^\star/(\lambda+1)$ becomes better off, implying that $\partial w_t(p)/\partial H_t^* \leq 0$ for all $p \leq H_t^*$. If this holds true, then we can say that poverty is unambiguously higher. Note that the function φ is continuous at $k^\star/(\lambda+1)$. Then it is readily verified that:

$$\frac{\partial \mu_{t+1}}{\partial H_t^*} = \beta\int_0^{H_t^*} [h'((\lambda+1)w_t(p))(\lambda+1) - \lambda r]\frac{\partial w_t(p)}{\partial H_t^*}\mathrm{d}p + \beta r\int_0^{H_t^*} \frac{\partial w_t(p)}{\partial H_t^*}\mathrm{d}p \tag{23.5}$$

The sign of this expression cannot be determined under the assumptions so far. It may be noted that, if there is (unrestricted) first-order dominance, whereby $\partial w_t(p)/\partial H_t^* \leq 0$ for all $p \in [0,1]$, then $\partial \mu_{t+1}/\partial H_t^* \leq 0$. However, first-order dominance is ruled out by the fact that the mean is held constant; there is a redistribution from the "wealth poor" to the "wealth nonpoor." On imposing a constant initial mean, $\mu_t = \overline{\mu}$, Equation (23.5) simplifies to:

$$\left[\frac{\partial \mu_{t+1}}{\partial H_t^*}\right]_{\mu_t=\overline{\mu}} = \beta \int_0^{H_t^*} [h'((\lambda+1)w_t(p)) - r](\lambda+1)\frac{\partial w_t(p)}{\partial H_t^*} \mathrm{d}p < 0 \qquad (23.6)$$

Thus, we find that an unambiguously higher initial headcount index of poverty holding the initial mean constant implies a lower growth rate, as claimed in Section 22.8.

REFERENCES

Abel-Smith, B., Townsend, P., 1966. The Poor and the Poorest, a New Analysis of the Ministry of Labour's Family Expenditure Surveys of 1953–54 and 1960. Bell, London.

Acemoglu, D., Robinson, J., 2006. Economic Origins of Dictatorship and Democracy. Cambridge University Press, Cambridge, England.

Acemoglu, D., Zilibotti, F., 1997. Was Prometheus unbound by chance? J. Polit. Econ. 105 (4), 709–751.

Adelman, I., Morris, C.T., 1973. Economic Growth and Social Equity in Developing Countries. Stanford University Press, Stanford.

Agee, J., Evans, W., 1941. Let Us Now Praise Famous Men, the American Classic, in Words and Pictures, of Three Tenant Families in the Deep South, Houghton Mifflin edition, Boston, MA, 2000.

Aghion, P., Bolton, P., 1997. A theory of trickle-down growth and development. Rev. Econ. Stud. 64, 151–172.

Aghion, P., Caroli, E., Garcia-Penalosa, C., 1999a. Inequality and economic growth, the perspectives of the new growth theories. J. Econ. Lit. 37 (4), 1615–1660.

Aghion, P., Banerjee, A., Piketty, T., 1999b. Dualism and macroeconomic volatility. Quart. J. Econ. 114 (4), 1359–1397.

Ahluwalia, M.S., 1976. Inequality, poverty and development. J. Dev. Econ. 3, 307–342.

Ahluwalia, M.S., Carter, N.G., Chenery, H.B., 1979. Growth and poverty in developing countries. J. Dev. Econ. 6, 299–341.

Ahluwalia, M.S., Duloy, J., Pyatt, G., Srinivasan, T.N., 1980. Who benefits from economic development? Comment. Am. Econ. Rev. 70 (1), 242–245.

Akerlof, G., 1970. The market for lemons, quality uncertainty and the market mechanism. Quart. J. Econ. 84, 485–500.

Akresh, R., de Walque, D., Kazianga, H., 2013. Cash Transfers and Child Schooling, Evidence from a Randomized Evaluation of the Role of Conditionality, Policy Research Working Paper 6340, World Bank.

Alatas, V., Banerjee, A., Hanna, R., Olken, B.A., Tobias, J., 2012. Targeting the poor, evidence from a field experiment in Indonesia. Am. Econ. Rev. 102 (4), 1206–1240.

Albelda, R., Folbre, N., The Center for Popular Economics, 1996. The War on the Poor. A Defense Manual. The New Press, New York.

Alderman, H., 2002. Do local officials know something we don't? Decentralization of targeted transfers in Albania. J. Pub. Econ. 83, 375–404.

Alderman, H., Hoddinott, J., Kinsey, B., 2006. Long-term consequences of early childhood malnutrition. Oxf. Econ. Pap. 58 (3), 450–474.

Alesina, A., Rodrik, D., 1994. Distributive politics and economic growth. Quart. J. Econ. 108, 465–490.

Allen, R., 2007. Pessimism Preserved, Real Wages in the British Industrial Revolution, Oxford University Department of Economics Working Paper 314.
Allen, R., 2009. Engels' pause technical change, capital accumulation, and inequality in the British industrial revolution. Explor. Econ. Hist. 46, 418–435.
Allen, R., 2013. Poverty Lines in History, Theory, and Current International Practice. Nuffield College, Oxford, Mimeo.
Anand, S., 1983. Inequality and Poverty in Malaysia. Oxford University Press, Oxford.
Anand, S., Kanbur, R., 1993. The Kuznets process and the inequality-development relationship. J. Dev. Econ. 40, 25–52.
Anand, S., Ravallion, M., 1993. Human development in poor countries, on the role of private incomes and public services. J. Econ. Perspect. 7 (Winter), 133–150.
Angelucci, M., Karlan, D., Zinman, J., 2013. Win Some Lose Some? Evidence from a Randomized Micro-credit Program Placement Experiment by Compartamos Banco. mimeo.
Aristotle, 350BC, Politics Book 2 (B. Jowett, Trans.). http://classics.mit.edu/Aristotle/politics.2.two.html.
Arneson, R., 1989. Equality and equal opportunity for welfare. Philos. Stud. 56 (1), 77–93.
Arrow, K.J., 1951. Social Choice and Individual Values. John Wiley, New York.
Artz, F.B., 1934. Reaction and Revolution, 1814–1834. Harper and Row, New York.
Atkinson, A.B., 1970. On the measurement of inequality. J. Econ. Theory 2, 244–263.
Atkinson, A.B., 1987. On the measurement of poverty. Econometrica 55, 749–764.
Atkinson, A.B., 1991. Comparing poverty rates internationally, lessons from recent studies in developed countries. World Bank Econ. Rev. 5 (1), 3–21.
Atkinson, A.B., Bourguignon, F., 1982. The comparison of multi-dimensional distributions of economic status. Rev. Econ. Stud. 49, 183–201.
Atkinson, A.B., Sutherland, H., 1989. Analysis of a partial basic income scheme. In: Atkinson, A.B. (Ed.), Poverty and Social Security. Harvester Wheatsheaf, Hertfordshire.
Attanasio, O., Meghir, C., Santiago, A., 2004. Education Choices in Mexico, Using a Structural Model and a Randomized Experiment to Evaluate PROGRESA. Working Paper EWP04/04, Institute of Fiscal Studies, London.
Aykroyd, W.R., 1974. The Conquest of Famine. Chatto and Windus, London.
Azariadis, C., 1996. The economics of poverty traps. Part one: complete markets. J. Econ. Growth 1, 449–486.
Azariadis, C., 2006. The theory of poverty traps, what have we learned? In: Bowles, S., Durlauf, S., Hoff, K. (Eds.), Poverty Traps. Princeton University Press, Princeton.
Baird, S., McIntosh, C., Ozler, B., 2011. Cash or condition? evidence from a cash transfer experiment. Quart. J. Econ. 126 (4), 1709–1753.
Banerjee, A., Duflo, E., 2003. Inequality and growth, what can the data Say? J. Econ. Growth 8 (3), 267–299.
Banerjee, A., Duflo, E., 2007. The economic lives of the poor. J. Econ. Perspect. 21 (2), 141–167.
Banerjee, A., Iyer, L., 2005. History, institutions and economic performance, the legacy of colonial land tenure systems in India. Am. Econ. Rev. 95 (4), 1190–1213.
Banerjee, A., Newman, A., 1993. Occupational choice and the process of development. J. Polit. Econ. 101 (2), 274–298.
Banerjee, Abhijit, Duflo, Esther, Glennerster, Rachel, Kinnan, Cynthia, 2009. The Miracle of Microfinance? Evidence from a Randomized Evaluation, BREAD Working Paper No. 278, BREAD.
Banerjee, A., Banerji, R., Duflo, E., Glennerster, R., Khemani, S., 2010. Pitfalls of participatory programs, evidence from a randomized evaluation in education in India. Am. Econ. J. Econ. Policy 2 (1), 1–30.
Bardhan, P., 1984a. Land, Labor and Rural Poverty, Essays in Development Economics. Columbia University Press, New York.
Bardhan, P., 1984b. The Political Economy of Development in India. Basil Blackwell, Oxford.
Bardhan, P., 2011. Challenges for a minimum social democracy in India. Econ. Pol. Wkly. 46 (10), 39–43.
Bardhan, P., Bowles, S., Ginitis, H., 2000. Wealth inequality, wealth constraints and economic performance. In: Atkinson, A.B., Bourguignon, F. (Eds.), In: Handbook of Income Distribution, vol. 1. North-Holland, Amsterdam.

Barrett, C.B., Marenya, P.P., McPeak, J.G., Minten, B., Murithi, F.M., Oluoch-Kosura, W., Place, F., Randrianarisoa, J.C., Rasambainarivo, J., Wangila, J., 2006. Welfare dynamics in rural Kenya and Madagascar. J. Dev. Stud. 42 (2), 178–199.
Barro, R., 2000. Inequality and growth in a panel of countries. J. Econ. Growth 5 (1), 5–32.
Barro, R., Sala-i-Martin, X., 1992. Convergence. J. Polit. Econ. 100 (2), 223–251.
Basu, K., 1999. Child labor, cause, consequence and cure, with remarks on international labor standards. J. Econ. Lit. 37, 1083–1119.
Bator, F.M., 1958. The anatomy of market failure. Quart. J. Econ. 72 (3), 351–379.
Beaudoin, S., 2007. Poverty in World History. Routledge, New York.
Beaumarchais, P., 1778. La Folle Journée ou Le Mariage de Figaro. Nathan, Paris, 2007.
Beck, T., Levine, R., Loayza, N., 2000. Finance and the sources of growth. J. Financ. Econ. 58, 261–300.
Beck, T., Demirguc-Kunt, A., Levine, R., 2007. Finance, inequality and the poor. J. Econ. Growth 12, 27–49.
Bénabou, R., 1996. Inequality and growth. In: Bernanke, B., Rotemberg, J. (Eds.), National Bureau of Economic Research Macroeconomics Annual. MIT Press, Cambridge.
Benton, D., 2010. The influence of dietary status on the cognitive performance of children. Mol. Nutr. Food Res. 54 (4), 457–470.
Berg, A., Ostry, J.D., Zettelmeyer, J., 2012. What makes growth sustained? J. Dev. Econ. 98, 149–166.
Besley, T., Burgess, R., 2003. The political economy of government responsiveness, theory and evidence from India. Quart. J. Econ. 117 (4), 1415–1451.
Bethlehem, J., 2009. The Rise of Survey Sampling, Statistics Netherlands Discussion Paper 09015, Amsterdam.
Beveridge, W., 1942. Social Insurance and Allied Services. His Majesty's Stationary Office, London.
Birdsall, N., Londono, J.L., 1997. Asset inequality matters, an assessment of the world Bank's approach to poverty reduction. Am. Econ. Rev. 87 (2), 32–37.
Birdsall, N., Ross, D., Sabot, R., 1995. Inequality and growth reconsidered, lessons from east Asia. World Bank Econ. Rev. 9 (3), 477–508.
Birdsall, N., Graham, C., Pettinato, S., 2000. Stuck in the Tunnel, Is Globalization Muddling the Middle Class? Center on Social and Economic Dynamics, Working Paper 14, Brookings Institution, Washington, DC.
Blaug, M., 1962. Economic Theory in Retrospect. Heinemann Books, London.
Blaug, M., 2001. No history of ideas, please, we're economists. J. Econ. Perspect. 15 (1), 145–164.
Bliss, C., 1975. Capital Theory and the Distribution of Income. North-Holland, Amsterdam.
Bliss, C., Stern, N., 1982. Palanpur, The Economy of an Indian Village. Clarendon Press, Oxford.
Boo, K., 2012. Behind the Beautiful Forevers. Random House, New York.
Booth, C., 1903. Life and labour of the people of London. Second Series: Industry, vol. 5. Macmillan and Co., London.
Bourguignon, F., 2003. The growth elasticity of poverty reduction, explaining heterogeneity across countries and time periods. In: Eicher, T., Turnovsky, S. (Eds.), Inequality and Growth Theory and Policy Implications. MIT Press, Cambridge.
Bourguignon, F., Morrisson, C., 2002. Inequality among world citizens, 1820–1992. Am. Econ. Rev. 92 (4), 727–744.
Bowles, S., Gintis, H., 1976. Schooling in Capitalist America, Educational Reform and the Contradictions of Economic Life. Routledge and Kegan Paul, London.
Bowles, S., Durlauf, S., Hoff, K., 2006. Poverty Traps. Princeton University Press, Princeton.
Bowley, A.L., 1915. The Nature and Purpose of the Measurement of Social Phenomena. P.S. King and Sons, London.
Boyer, G., 2002. English poor laws. In: Whaples, R. (Ed.), EH.Net Encyclopedia.
Brinton, C., 1934. A Decade of Revolution 1789–1799. Harper and Row, New York.
Brooke, M.Z., 1998. Le Play Engineer and Social Scientist. Transaction Publishers, New Brunswick.
Bruno, M., Ravallion, M., Squire, L., 1998. Equity and growth in developing countries, Old and New perspectives on the policy issues. In: Tanzi, V., Chu, K.-y. (Eds.), Income Distribution and High-Quality Growth. MIT Press, Cambridge.

Bulow, J., Summers, L., 1986. A theory of dual labor markets with application to industrial policy, discrimination and Keynesian unemployment. J. Labor Econ. 4 (3), 376–414.

Burnett, J., 1969. A History of the Cost of Living. Penguin Books, Harmondsworth, England.

Cameron, L., 2002. Did Social Safety Net Scholarships Reduce Drop-Out Rates during the Indonesian Economic Crisis? Policy Research Working Paper 2800, World Bank.

Campbell, F., Conti, G., Heckman, J., Moon, S.H., Pinto, R., Pungello, E., Pan, Y., 2014. Early childhood investments substantially boost adult health. Science 343, 1478–1485.

Cantril, H., 1965. The Pattern of Human Concerns. Rutgers University Press, New Brunswick.

Cappelen, A., Mæstad, O., Tungodden, B., 2010. Demand for childhood vaccination-insights from behavioral economics. Forum Dev. Stud. 37 (3), 349–364.

Carter, M., Barrett, C., 2006. The economics of poverty traps and persistent poverty, an asset-based approach. J. Dev. Stud. 42 (2), 178–199.

Chen, S., Ravallion, M., 2010. The developing world is poorer than we thought, but no less successful in the fight against poverty. Quart. J. Econ. 125 (4), 1577–1625.

Chen, S., Ravallion, M., 2013. More relatively poor people in a less absolutely poor world. Rev. Income Wealth 59 (1), 1–28.

Chen, S., Ren, M., Ravallion, M., 2009. Are there lasting impacts of aid to poor areas? J. Pub. Econ. 93, 512–528.

Chenery, H., 1977. Forward to David Morawetz, Twenty-Five Years of Economic Development. World Bank, Washington, DC.

Chenery, H., Ahluwalia, M.S., Bell, C., Duloy, J., Jolly, R., 1974. Redistribution with Growth. Oxford University Press, Oxford.

Cinnirella, F., 2008. Optimists or Pessimists? A Reconsideration of Nutritional Status in Britain, 1740–1865. Eur. Rev. Econ. Hist. 12, 325–354.

Clark, G., 2005. The condition of the working class in England 1209–2004. J. Polit. Econ. 113, 1307–1340.

Clarke, G.R.G., 1995. More evidence on income distribution and growth. J. Dev. Econ. 47, 403–428.

Cogneau, D., 2012. The Political Dimension of Inequality during Economic Development. Région et Développement 35, 11–36.

Cohen, G., 1989a. On the currency of egalitarian justice. Ethics 99 (4), 906–944.

Cohen, J., 1989b. Democratic equality. Ethics 99 (4), 727–751.

Cornia, Giovanni, Stewart, Frances, 1995. Two errors of targeting. In: van de Walle, Dominique, Nead, Kimberly (Eds.), Public Spending and the Poor. Johns Hopkins University Press for the World Bank, Baltimore.

Cunha, F., Heckman, J., 2007. The technology of skill formation. Am. Econ. Rev. 97 (2), 31–47.

Cunningham, H., 1990. The employment and unemployment of children in England c.1680-1851. Past Present 126, 115–150.

Currie, J., 2001. Early childhood development programs. J. Econ. Perspect. 15 (2), 213–238.

Currie, J., 2011. Inequality at birth, some causes and consequences. Am. Econ. Rev. 101 (3), 1–22.

Currie, J., 2012. Antipoverty programs for poor children and families. In: Jefferson, P.N. (Ed.), The Oxford Handbook of the Economics of Poverty. Oxford University Press, Oxford.

Dandekar, V.M., Rath, N., 1971. Poverty in India. Indian School of Political Economy, Pune.

Daniels, W., 1898. The bearing of the doctrine of selection upon the social problem. Inter. J. Ethics 8 (2), 203–214 (now called Ethics).

Das, J., Do, Q.-T., Ozler, B., 2005. A welfare analysis of conditional cash transfer schemes. World Bank Res. Obs. 20 (1), 57–80.

Dasgupta, P., 1993. An Inquiry into Well-Being and Destitution. Oxford University Press, Oxford.

Dasgupta, P., 1997. Poverty traps. In: Kreps, D.M., Wallis, K.F. (Eds.), Advances in Economics and Econometrics, Theory and Applications. Cambridge University Press, Cambridge.

Dasgupta, P., 2011. Personal Histories and Poverty Traps. In: Annual World Bank Conference on Development Economics. World Bank, Washington, DC.

Dasgupta, P., Ray, D., 1986. Inequality as a determinant of malnutrition and unemployment. Econ. J. 96, 1011–1034.

Datt, G., Ravallion, M., 1998. Farm productivity and rural poverty in India. J. Dev. Stud. 34 (4), 62–85.

Day, R.H., 1992. Complex economic dynamics, obvious in history, generic in theory, elusive in data. J. Appl. Econ. 7, S9–S23.
De Donder, P., Hindriks, J., 1998. The political economy of targeting. Public Choice 95, 177–200.
De Janvry, A., Sadoulet, E., 2006. Making conditional cash transfer programs more efficient, designing for maximum effect of the conditonality. World Bank Econ. Rev. 20 (1), 1–29.
De Mandeville, B., 1732. An essay on charity and charity schools. In: The Fable of the Bees, Or, Private Vices, Publick Benefits, sixth ed. Oxford University Press, London, Oxford, 1957 (reprint).
de Montesquieu, C., 1748. The Spirit of Laws (T. Nugent, Trans.). G. Bell and Sons, London (1914 edition).
De Tocqueville, A., 1835. Memoir on Pauperism, Does Public Charity Produce and Idle and Dependent Class of Society? Cosimo Classics edition, New York, 2005.
de Waal, F., 2009. The Age of Empathy. Nature's Lessons for a Kinder Society. Three Rivers Press, New York.
Dearden, L., Emmerson, C., Frayne, C., Meghir, C., 2009. Conditional cash transfers and school dropout rates. J. Hum. Res. 44 (4), 827–857.
Deaton, A., 1997. The Analysis of Household Surveys: A Microeconometric Approach to Development Policy. World Bank, Washington, DC.
Deaton, A., 2006. The great escape, a review of Robert Fogel's the escape from hunger and premature death, 1700–2100. J. Econ. Lit. 44, 106–114.
Deininger, K., Olinto, P., 2000. Asset Distribution, Inequality and Growth, Policy Research Working Paper 2375, World Bank.
Deininger, K., Squire, L., 1996. A New data set measuring income inequality. World Bank Econ. Rev. 10, 565–591.
Deininger, K., Squire, L., 1998. New ways of looking at old issues, inequality and growth. J. Dev. Econ. 57 (2), 259–287.
Dercon, S., Outes, I., 2013. The Road to Perdition, Rainfall Shocks, Poverty Traps and Destitution in Semi-Arid India. Oxford University, mimeo.
Dickens, C., 1838. Oliver Twist. Richard Bently, London, Penguin edition, 2003.
Do, Q.-T., Levchenko, A., Ravallion, M., 2013. Copying with food price volatility, trade insulation as social protection. In: Chavas, J.-P., Hummels, D., Wright, B. (Eds.), The Economics of Food Price Volatility. University of Chicago Press, forthcoming.
Doepke, M., Zilibotti, F., 2005. Social class and the spirit of capitalism. J. Eur. Econ. Assoc. 3 (2–3), 516–524.
Doeringer, P., Piore, M., 1971. Internal Labor Markets and Manpower Analysis. Sharpe, New York.
Dollar, D., Kraay, A., 2002. Growth is good for the poor. J. Econ. Growth 7 (3), 195–225.
Dollar, D., Kraay, A., 2004. Trade, growth and poverty. Econ. J. 114 (493), F22–F49.
Doron, A., 1990. Definition and measurement of poverty—the unsolved issue. In: Social Security, Journal of Welfare and Social Security Studies, 2, pp. 27–50, Special English Edition.
Drèze, J., 1990a. Famine prevention in India. In: Drèze, Sen (Eds.), In: The Political Economy of Hunger, vol. 2. Oxford University Press, Oxford.
Drèze, J., 1990b. Famine prevention in Africa, some experiences and lessons. In: Drèze, J., Sen, A. (Eds.), The Political Economy of Hunger. vol. 2. Oxford University Press, Oxford.
Drèze, J., Sen, A., 1989. Hunger and Public Action. Oxford University Press, Oxford.
Duncan, O., Featherman, D., Duncan, B., 1972. Socioeconomic Background and Achievement. Seminar Press, New York.
Dutta, P., Murgai, R., Ravallion, M., van de Walle, D., 2012. Does India's employment guarantee scheme guarantee employment? Econ. Pol. Wkly. 48 (April 21), 55–64.
Dutta, P., Murgai, R., Ravallion, M., van de Walle, D., 2013. Right-to-Work? Assessing India's Employment Guarantee Scheme in Bihar. World Bank, Washington, DC.
Easterly, W., 2001. The middle class consensus and economic development. J. Econ. Growth 6 (4), 317–335.
Easterly, W., 2009. How the millennium development goals are unfair to Africa. World Dev. 37 (1), 26–35.
Eastwood, R., Lipton, M., 1999. The impact of changes in human fertility on poverty. J. Dev. Stud. 36 (1), 1–30.

Eastwood, R., Lipton, M., 2001. Demographic transition and poverty, effects via economic growth, distribution and conversion. In: Birdsall, N., Kelley, A., Sinding, S. (Eds.), Population Matters. Oxford University Press, Oxford.
Eden, Frederick Morton, 1797. The State of the Poor. J. Davis, London.
Elbers, C., Lanjouw, J., Lanjouw, P., 2003. Micro-level estimation of poverty and inequality. Econometrica 71 (1), 355–364.
Ellwood, D., Summers, L., 1986. Poverty in america, is welfare the answer or the problem? In: Danziger, S., Weinberg, D. (Eds.), Fighting Poverty, What Works and What Doesn't. Harvard University Press, Cambridge.
Engel, E., 1857. Die Productions- und Consumtionsverhältnisse des Königreichs Sachsen. Zeitschrift des statistischen Bureaus des Königlich Sächsischen Ministerium des Inneren 8–9, 28–29.
Engels, F., 1845. The Condition of the Working Class in England. Oxford University Press, Oxford, 1993 edition.
Engerman, S.L., Sokoloff, K., 2006. Colonialism, inequality and long-Run paths of development. In: Banerjee, A., Benabou, R., Mookherjee, D. (Eds.), Understanding Poverty. Oxford University Press, Oxford, pp. 37–62.
Eurostat, 2005. Income poverty and social exclusion in the EU25, Statistics in Focus 03/2005, Office of Official Publications of the European Communities, Luxembourg.
Ferreira, F.H.G., Ravallion, M., 2009. Poverty and inequality, the global context. In: Salverda, W., Nolan, B., Smeeding, T. (Eds.), The Oxford Handbook of Economic Inequality. Oxford University Press, Oxford.
Fields, G.S., 1977. Who benefits from economic development? a reexamination of Brazilian growth in the 196O's. Am. Econ. Rev. 67 (4), 570–582.
Fields, G.S., 1980. Poverty Inequality and Development. Cambridge University Press, Cambridge.
Fields, G.S., 2001. Distribution and Development. Russell Sage Foundation, New York.
Fifield, A., 2013. Starved of healthy options. Financ. Times, 9, June 14.
Filmer, D., 2007. If You build It, will they come? school availability and school enrolment in 21 poor countries. J. Dev. Stud. 43 (5), 901–928.
Fishlow, A., 1972. Brazilian size distribution of income. Am. Econ. Rev. 62, 391–402.
Fishlow, A., Gwin, C., 1994. Overview, Lessons from the East Asian Experience. In: Fishlow, A., Gwin, C., Haggard, S., Rodrik, D., Wade, R. (Eds.), Miracle or Design? Lessons from the East Asian Experience. Overseas Development Council, Washington, DC.
Fiszbein, A., Schady, N., 2010. Conditional Cash Transfers for Attacking Present and Future Poverty. World Bank, Washington, DC.
Fleischacker, S., 2004. A Short History of Distributive Justice. Harvard University Press, Cambridge, MA.
Fleurbaey, M., 2008. Fairness Responsibility and Welfare. Oxford University Press, Oxford.
Fogel, R.W., 2004. The Escape from Hunger and Premature Death, 1700–2100. Cambridge University Press, Cambridge.
Fogel, R.W., Engerman, S., Floud, R., Friedman, G., Mango, R., Sokoloff, K., Steckel, R., Trussell, J., Villaflor, G., Watchter, K., 1983. Secular change in American and British stature and nutrition. In: Rotberg, R., Rabb, T. (Eds.), Hunger in History. Cambridge University Press, Cambridge.
Forbes, K.J., 2000. A reassessment of the relationship between inequality and growth. Am. Econ. Rev. 90 (4), 869–887.
Foster, J., 1998. Absolute versus relative poverty. Am. Econ. Rev. 88 (2), 335–341.
Foster, J., Greer, J., Thorbecke, E., 1984. A class of decomposable poverty measures. Econometrica 52, 761–765.
Friedman, M., 1962. Capital and Freedom. University of Chicago Press, Chicago.
Fuchs, V., 1967. Redefining poverty and redistributing income. Public Interest 8, 88–95.
Furniss, E., 1920. The Position of the Laborer in a System of Nationalism. A Study in the Labor Theories of the Later English Mercantilists. Houghton Mifflin, Boston and New York.
Gaiha, R., 1997. Rural public works and the poor, the case of the employment guarantee scheme in India. In: Polachek, S. (Ed.), Research in Labour Economics. JAI Press, Connecticut.
Galasso, E., Ravallion, M., 2005. Decentralized targeting of an antipoverty program. J. Pub. Econ. 89, 705–727.

Galasso, E., Ravallion, M., Salvia, A., 2004. Assisting the transition from workfare to work: Argentina's Proempleo Experiment. Ind. Labor Relat. Rev. 57 (5), 128–142.
Galbraith, J.K., 1958. The Affluent Society. Mariner Books, Boston.
Galor, O., Zeira, J., 1993. Income distribution and macroeconomics. Rev. Econ. Stud. 60 (1), 35–52.
Gans, H., 1995. The War Against the Poor. Basic Books, New York.
Garces, E., Thomas, D., Currie, J., 2002. Longer term effects of head start. Am. Econ. Rev. 92 (4), 999–1012.
Gelbach, J., Pritchett, L., 2000. Indicator targeting in a political economy, leakier can be better. J. Policy Reform 4, 113–145.
Geremek, B., 1994. Poverty. A History. Blackwell, Oxford.
Ghatak, M., Jiang, N.N.-H., 2002. A simple model of inequality, occupational choice, and development. J. Dev. Econ. 69 (1), 205–226.
Gillie, A., 1996. The origin of the poverty line. Econ. Hist. Rev. 49 (4), 715–730.
Goldin, C., Katz, L.F., 2008. The Race Between Education and Technology. Harvard University Press, Cambridge.
Gordon, R., Li, W., 2009. Tax structures in developing countries, many puzzles and a possible explanation. J. Pub. Econ. 93 (7–8), 855–866.
Grosh, M., Glewwe, P., 2000. Designing Household Survey Questionnaires for Developing Countries. Lessons from 15 Years of the Living Standards Measurement Study, 3 vols. World Bank, Washington, DC.
Grosh, M., del Ninno, C., Tesliuc, E., Ouerghi, A., 2008. For Protection and Promotion, the Design and Implementation of Effective Safety Nets. World Bank, Washington, DC.
Gutiérrez, C., Tanaka, R., 2009. Inequality and education decisions in developing countries. J. Econ. Inequal. 7, 55–81.
Hansen, M.H., Hurvitz, W.N., 1943. On the theory of sampling from a finite population. Ann. Math. Stat. 14, 333–362.
Harrington, M., 1962. The Other America, Poverty in the United States. Macmillan, New York.
Harrison, A., Rodríguez-Clare, A., 2010. Trade, foreign investment, and industrial policy for developing countries. In: Handbook of Development Economics. vol. 5. North Holland, Amsterdam.
Harsanyi, J., 1975. Can the maximin principle serve as a basis for morality? A critique of john Rawls's theory. Am. Polit. Sci. Rev. 69 (2), 594–606.
Hartmann, B., 1987. Reproductive Rights and Wrongs, the Global Politics of Population Control and Contraceptive Choice. Harper and Row, New York.
Hauk, W.R., Wacziarg, R., 2009. A Monte Carlo study of growth regressions. J. Econ. Growth 14 (2), 103–147.
Hazan, M., Berdugo, B., 2002. Child labour, fertility, and economic growth. Econ. J. 112 (482), 810–828.
Heckman, J., 2006. Skill formation and the economics of investing in disadvantaged children. Science 30, 1900–1902.
Heclo, H., 1986. The political foundations of antipoverty policy. In: Danziger, S., Weinberg, D. (Eds.), Fighting Poverty, What Works and What Doesn't. Harvard University Press, Cambridge.
Hemerijck, A., 2014. The reform capacities of European welfare states. In: Cantillon, B., Vandenbroucke, F. (Eds.), Reconciling Work and Poverty Reduction. How Successful are European Welfare States? Oxford University Press, Oxford.
Hertel-Fernandez, A., Wenger, J.B., 2013. Taking Up Social Benefits, A Cautionary Tale from an Unemployment Insurance Survey Experiment, Available at SSRN, http://ssrn.com/abstract=2341885 or http://dx.doi.org/10.2139/ssrn.2341885.
Herzer, D., Vollmer, S., 2012. Inequality and growth, evidence from panel cointegration. J. Econ. Inequal. 10, 489–503.
Hill, C., 1972. The World Turned Upside Down, Radical Ideas During the English Revolution. Maurice Temple Smith, London.
Himmelfarb, G., 1984a. The Idea of Poverty, England in the Early Industrial Age. Faber and Faber, London.
Himmelfarb, G., 1984b. The idea of poverty. Hist. Today 34 (4). http://www.historytoday.com/gertrude-himmelfarb/idea-poverty.

Hindle, S., 2004. On the Parish? The Micro-Politics of Poor Relief in Rural England 1550–1750. Oxford University Press, Oxford.
Hoff, K., 1996. Market failures and the distribution of wealth, a perspective from the economics of information. Polit. Soc. 24 (4), 411–432.
Hossain, M., 1988. Credit for Alleviation of Rural Poverty, the Grameen Bank in Bangladesh: IFPRI Research Report 65. International Food Policy Research Institute, Washington, DC.
Howe, I., 1993. Introduction to Michael Harrington's, the Other America. 1993 Reprinting, Touchstone, New York.
Hoynes, H., 1997. Does welfare play any role in female headship decisions? J. Publ. Econ. 65, 89–117.
Hunter, R., 1904. Poverty. MacMillan Company, London.
Imbert, C., Papp, J., 2011. Estimating leakages in India's employment guarantee. In: Khera, R. (Ed.), The Battle for Employment Guarantee. Oxford University Press, New Delhi, pp. 269–278.
Jalan, J., Ravallion, M., 1998. Are there dynamic gains from a poor-area development program? J. Pub. Econ. 67 (1), 65–86.
Jalan, J., Ravallion, M., 2002. Geographic poverty traps? A micro model of consumption growth in rural China. J. Appl. Econ. 17 (4), 329–346.
Jalan, J., Ravallion, M., 2004. Household income dynamics in rural china. In: Dercon, S. (Ed.), Insurance Against Poverty. Oxford University Press, New York, pp. 107–123.
Jha, R., Gaiha, R., Pandey, M.K., 2012. Net transfer benefits under India's rural employment guarantee scheme. J. Policy Model 34 (2), 296–311.
Johnson, D., Smeeding, T., 2012. A Consumer's Guide to Interpreting Various U.S. Poverty Measures, Fast Focus No. 14, Institute for Research on Poverty, University of Wisconsin.
Jones, G.S., 2004. An End to Poverty? A Historical Debate. Columbia University Press, New York.
Just, D., Michelson, H., 2007. Wealth as welfare, Are wealth thresholds behind persistent poverty? Appl. Econ. Perspect. Policy 29 (3), 419–426.
Kahl, S., 2005. The religious roots of modern poverty policy, catholic, Lutheran, and reformed protestant traditions compared. Eur. J. Sociol. 46, 91–126.
Kakwani, N., 1980. Income Inequality and Poverty, Methods of Estimation and Policy Applications. Oxford University Press, Oxford.
Kaldor, N., 1955. Alternative theories of distribution. Rev. Econ. Stud. 23 (2), 94–100.
Kalecki, M., 1942. A theory of profits. Econ. J. 52, 258–267.
Kanbur, R., 1987. Measurement and alleviation of poverty. IMF Staff. Pap. 36, 60–85.
Kanbur, R., Tuomala, M., 2011. Charitable conservatism, poverty radicalism and inequality aversion. J. Econ. Inequal. 9, 417–431.
Kanbur, R., Keen, M., Tuomala, M., 1994. Labor supply and targeting in poverty alleviation programs. World Bank Econ. Rev. 8 (2), 191–211.
Kant, I., 1785. Fundamental Principles of the Metaphysic of Morals. In: Kingsmill Abbott, T. (Ed.), 10 ed, Project Gutenberg.
Katz, M.B., 1987. The Undeserving Poor, From the War on Poverty to the War on Welfare. Pantheon Books, New York.
Keefer, P., Khemani, S., 2005. Democracy, Public Expenditures, and the Poor, Understanding Political Incentives for Providing Public Services. World Bank Res. Obs. 20 (1), 1–28.
Kelley, A., Schmidt, R., 1995. Aggregate population and economic growth correlations, the role of the components of demographic change. Demography 32 (4), 543–555.
Kelley, A., Schmidt, R., 2001. Economic and demographic change, a synthesis of models, findings and perspectives. In: Birdsall, N., Kelley, A., Sinding, S. (Eds.), Population Matters. Oxford University Press, Oxford.
Kelly, M., Cormac, Ó.G., 2010. Living Standards and Mortality since the Middle Ages, Working Paper 201026, School Of Economics, University College Dublin.
Keynes, J.M., 1936. The General Theory of Employment, Interest and Money. Macmillan Press, London.
Keysers, C., 2011. Empathic Brain. How the Discovery of Mirror Neurons Changes our Understanding of Human. Social Brain Press, Nature.

Khandker, S., Samad, H., 2014. Dynamic Effects of Microcredit in Bangladesh. Policy Research Working Paper 6821, World Bank, Washington, DC.
Kinsey, B., 2013. The excluded generations, questioning a leading poverty indicator. In: Paper presented at the UNU-WIDER conference "Inclusive Growth in Africa, Measurement, Causes, and Consequences.
Klebaner, B.J., 1964. Poverty and its relief in American thought, 1815–61. Soc. Serv. Rev. 38 (4), 382–399.
Knowles, S., 2005. Inequality and economic growth, the empirical relationship reconsidered in the light of comparable data. J. Dev. Stud. 41 (1), 135–159.
Kolm, S.-C., 1976. Unequal inequalities. I. J. Econ. Theory 12 (3), 416–442.
Korpe, P., Petri, W., 2012. Environmental enteropathy, critical implications of a poorly understood condition. Trends Mol. Med. 18 (6), 328–336.
Kraay, A., 2006. When is growth Pro-poor? Evidence from a panel of countries. J. Dev. Econ. 80, 198–227.
Kuznets, S., 1955. Economic growth and income inequality. Am. Econ. Rev. 45, 1–28.
Landauer, C., 1959. European Socialism, a History of Ideas and Movements from the Industrial Revolution to Hitler's Seizure of Power. University of California Press, Berkeley.
Lanjouw, P., Stern, N., 1998. Economic Development in Palanpur over Five Decades. Clarendon Press, Oxford.
Lepenies, P.H., 2014. Of goats and dogs, Joseph Townsend and the idealisation of markets—a decisive episode in the history of economics. Camb. J. Econ. 38, 447–457.
Levine, R., Renelt, D., 1992. A sensitivity analysis of cross-country growth regressions. Am. Econ. Rev. 82, 942–963.
Lewis, A., 1954. Economic development with unlimited supplies of labor. Manchester Sch. Econ. Soc. Stud. 22, 139–191.
Li, H., Zou, H.-f., 1998. Income inequality is not harmful to growth, theory and evidence. Rev. Dev. Econ. 2 (3), 318–334.
Lin, J.Y., 2012. New Structural Economics, a Framework for Rethinking Development and Policy. World Bank, Washington, DC.
Lindert, P.H., 2004. Growing Public: The Story, Social Spending and Economic Growth since the Eighteenth Century, vol. 1. Cambridge University Press, Cambridge.
Lindert, P.H., 2013. Private welfare and the welfare state. In: Neal, L., Williamson, J. (Eds.), The Cambridge History of Capitalism. Cambridge University Press, Cambridge.
Lipton, M., 1968. Urban bias and rural planning, strategy for agriculture. In: Streeten, P., Lipton, M. (Eds.), The Crisis in Indian Planning. Oxford University Press, Oxford.
Lipton, M., 1977. Why Poor People Stay Poor, Urban Bias and World Development. Temple Smith, London.
Lokshin, M., Ravallion, M., 2004. Household income dynamics in two transition economies. Stud. Nonlinear Dyn. Econom. 8 (3), 1–33.
Lopez, H., Servén, L., 2006. A Normal Relationship? Poverty, Growth and Inequality, Policy Research Working Paper 3814, World Bank.
Lopez, H., Servén, L., 2009. Too Poor to Grow, Policy Research Working Paper 5012, World Bank.
Loury, G., 1981. Intergenerational transfers and the distribution of earnings. Econometrica 49, 843–867.
Ludwig, J., Phillips, D.A., 2007. The Benefits and Costs of Head Start, NBER Working Paper 12973.
Lundberg, M., Squire, L., 2003. The simultaneous evolution of growth and inequality. Econ. J. 113, 326–344.
Macours, K., Schady, N., Vakis, R., 2008. Cash Transfers, Behavioral Changes and Cognitive Development in Early Childhood. Policy Research Working Paper 4759, World Bank, Washington, DC.
Mahalanobis, P.C., 1953. Some observations on the process of growth. Sankhya 12, 307–312.
Malthus, T.R., 1806. An Essay on the Principle of Population, 1890 ed. Ward, Lock and Co., London.
Mani, A., Mullainathan, S., Shafir, E., Zhao, J., 2013. Poverty impedes cognitive function. Science 341, 976–980.
Mansuri, G., Rao, V., 2012. Localizing Development, Does Participation Work? World Bank, Washington, DC.
Marshall, A., 1890. Principles of Economics. Macmillan, London (8th ed., 1920).
Marshall, A., 1907. Some possibilities of economic chivalry. Econ. J. 17 (65), 7–29.

Martin, W., Anderson, K., 2012. Export restrictions and price insulation during commodity price booms. Am. J. Agr. Econ. 94 (2), 422–427.
Marx, K., Engels, F., 1848. Manifesto of the Communist Party, London.
Mayhew, H., 2008. London Labour and the London Poor. Wordsworth Classics, London (reprinting a selection of newspaper articles from the 1840s).
McKenzie, D., Woodruff, C., 2006. Do entry costs provide an empirical basis for poverty traps? Evidence from Mexican microenterprises. Econ. Dev. Cult. Change 55 (1), 3–42.
McNamara, R., 1973. Address to the Board of Governors at the 1973 Annual General Meeting. World Bank, Washington DC.
Meade, J., 1972. Poverty in the welfare state. Oxf. Econ. Pap. 24, 289–326.
Mencher, S., 1967. Poor Law to Poverty Program, Economic Security Policy in Britain and the United States. University of Pittsburgh Press, Pittsburgh.
Mesnard, A., Ravallion, M., 2006. The wealth effect on new business startups in a developing economy. Economica 73, 367–392.
Meyer, B.D., Sullivan, J., 2012. Consumption and income poverty in the United States. In: Jefferson, P.N. (Ed.), The Oxford Handbook of the Economics of Poverty. Oxford University Press, Oxford.
Michel, J.-B., Shen, Y.K., Aiden, A.P., Veres, A., Gray, M.K., The Google Books Team, Pickett, J.P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M.A., Aiden, E.L., 2010. Quantitative analysis of culture using millions of digitized books. Science, 16 December.
Milanovic, B., 2005. Can we discern the effect of globalization on income distribution? World Bank Econ. Rev. 19 (1), 21–44.
Milanovic, B., 2008. Qat expenditures in Yemen and Djibouti, an empirical analysis. J. Afr. Econ. 17 (5), 661–687.
Mill, J.S., 1848. Principles of Political Economy, 1965 ed. A.M. Kelly, New York.
Mill, J.S., 1859. On Liberty, Dover Thrift Edition, Toronto2002.
Mirrlees, J., 1971. An exploration in the theory of optimum income taxation. Rev. Econ. Stud. 38, 175–208.
Moehling, C.M., 1999. State child labor laws and the decline of child labor. Explor. Econ. Hist. 36 (1), 72–106.
Montagu, A., 1971. Forward, in the 1971 reprint of Townsend (1786).
Morduch, J., 1999. The role of subsidies in microfinance, evidence from the Grameen Bank. J. Dev. Econ. 60, 229–248.
Muller, J.Z., 1993. Adam Smith in his Time and Ours, Designing a Decent Society. Princeton University Press, Princeton.
Murgai, R., Ravallion, M., 2005. Is a Guaranteed Living Wage a Good Antipoverty Policy? Policy Research Working Paper 3460, World Bank.
Murgai, R., Ravallion, M., van de Walle, D., 2013. Is Workfare Cost-Effective against Poverty in a Poor Labor-Surplus Economy? Policy Research Working Paper, World Bank.
Murphy, K., Schleifer, A., Vishny, R., 1989. Industrialization and the big push. J. Polit. Econ. 97 (5), 1003–1026.
Murray, C.A., 1984. Losing ground: American Social Policy 1950–1980. Basic Books, New York.
Musgrave, R., 1985. A brief history of fiscal doctrine. In: Auerbach, A.J., Feldstein, M. (Eds.), In: Handbook of Public Economics, vol. 1. North-Holland, Amsterdam.
Nath, S.K., 1969. A Reappraisal of Welfare Economics. Routledge and Kegan Paul, London.
Nozick, R., 1974. Anarchy, State and Utopia. Basic Books, New York.
O'Connor, A., 2002. Poverty Knowledge, Social Science, Social Policy, and the Poor in Twentieth-Century U.S. HistoryPrinceton University Press, Princeton.
O'Rourke, K., Williamson, J., 1997. Globalization and History, the Evolution of the Nineteenth Century Atlantic Economy. MIT Press, Cambridge, MA.
Orshansky, M., 1963. Children of the poor. Soc. Secur. Bull. 26, 3–29.
Pack, H., Saggi, K., 2006. The Case for Industrial Policy, a Critical Survey. Policy Research Working Paper 3839World Bank, Washington, DC.
Paine, T., 1797. Agrarian Justice, 2004 edition published with Common Sense by Penguin.

Pandey, P., Goyal, S., Sundararaman, V., 2009. Community participation in public schools, impacts of information campaigns in three indian states. Educ. Econ. 13 (3), 355–375.
Perotti, R., 1996. Growth, income distribution and democracy, what the data say. J. Econ. Growth 1 (2), 149–187.
Persson, T., Tabellini, G., 1994. Is inequality harmful for growth? Am. Econ. Rev. 84, 600–621.
Pigou, A., 1920. The Economics of Welfare, Fourth ed. Macmillan, London, 1971.
Pitt, M., Khandker, S., 1998. The impact of group-based credit programs on poor households in Bangladesh, does the gender of participants matter? J. Polit. Econ. 106, 958–996.
Pitt, M., Khandker, S., 2012. Replicating Replication Due Diligence in Roodman and Morduch's Replication of Pitt and Khandker (1998). Policy Research Working Paper 6273, World Bank, Washington, DC.
Pogge, T.W., 1989. Realizing Rawls. Cornell University Press, Ithaca.
Pollak, R., Wales, T., 1979. Welfare comparison and equivalence scale. Am. Econ. Rev. 69, 216–221.
Pradhan, M., Ravallion, M., 2000. Measuring poverty using qualitative perceptions of consumption adequacy. Rev. Econ. Stat. 82 (3), 462–471.
Probe Team, 1999. Public Report on Basic Education in India. Oxford University Press, New Delhi.
Proceedings of the Old Bailey., 2012. London, 1760–1815. Web Site, London.
Rajan, R., 2009a. Rent preservation and the persistence of underdevelopment. Am. Econ. J. Macroecon. 1 (1), 178–218.
Rajan, R., 2009b. Saving growth from unequal influence. In: Levy, S., Walton, M. (Eds.), No Growth without Equity? Inequality, Interests and Competition in Mexico. World Bank, Washington, DC.
Rao, V., 2001. Poverty and public celebrations in rural India. Ann. Am. Acad. Pol. Soc. Sci. 573 (1), 85–104.
Ravallion, M., 1987. Trade and stabilization: another look at British India's controversial foodgrain exports. Explor. Econ. Hist. 24, 354–370.
Ravallion, M., 1994. Measuring social welfare with and without poverty lines. Am. Econ. Rev. 84 (2), 359–365.
Ravallion, M., 1995. Growth and poverty, evidence for developing countries in the 1980s. Econ. Lett. 48, 411–417.
Ravallion, M., 1997a. Can high inequality developing countries escape absolute poverty? Econ. Lett. 56, 51–57.
Ravallion, M., 1997b. Famines and economics. J. Econ. Lit. 35 (3), 1205–1242.
Ravallion, M., 1998. Does aggregation hide the harmful effects of inequality on growth? Econ. Lett. 61 (1), 73–77.
Ravallion, M., 2001. Growth, inequality and poverty, looking beyond averages. World Dev. 29 (11), 1803–1815.
Ravallion, M., 2003. Inequality convergence. Econ. Lett. 80, 351–356.
Ravallion, M., 2005. Externalities in rural development, evidence for China. In: Kanbur, R., Venables, T. (Eds.), Spatial Inequality and Development. Oxford University Press, pp. 137–162.
Ravallion, M., 2006. Looking beyond averages in the trade and poverty Debate. World Dev. 34 (8), 1374–1392.
Ravallion, M., 2007. Inequality is bad for the poor. In: Micklewright, J., Jenkins, S. (Eds.), Inequality and Poverty Re-Examined. Oxford University Press, Oxford.
Ravallion, M., 2009. Are there lessons for Africa from China's success against poverty? World Dev. 37 (2), 303–313.
Ravallion, M., 2011. The two poverty enlightenments, historical insights from digitized books spanning three centuries. Poverty Public Policy 3 (2), 1–45.
Ravallion, M., 2012a. Poverty lines across the world. In: Jefferson, P.N. (Ed.), The Oxford Handbook of the Economics of Poverty. Oxford University Press, Oxford, pp. 75–104.
Ravallion, M., 2012b. Why don't we see poverty convergence? Am. Econ. Rev. 102 (1), 504–523.
Ravallion, M., 2013. How long will it take to lift one billion people out of poverty? World Bank Res. Obs. 28 (2), 139–158.
Ravallion, M., 2014. Poor, or just feeling poor. In: Clark, A., Senik, C. (Eds.), Subjective Wellbeing in Developing Countries. Oxford University Press, Oxford, pp. 140–174.

Ravallion, M., 2015. The Economics of Poverty: History, Measurement and Policy. Oxford University Press, New York.
Ravallion, M., Chen, S., 2007. China's (uneven) progress against poverty. J. Dev. Econ. 82 (1), 1–42.
Ravallion, M., Chen, S., 2011. Weakly relative poverty. Rev. Econ. Stat. 93 (4), 1251–1261.
Ravallion, M., Chen, S., 2013. A proposal for truly global poverty measures. Glob. Policy 4 (3), 258–265.
Ravallion, M., Datt, G., 1995. Is targeting through a work requirement efficient? Some evidence for rural India. In: de Walle, D., Nead, K. (Eds.), Public Spending and the Poor: Theory and Evidence. Johns Hopkins University Press, Baltimore.
Ravallion, M., Datt, G., 2002. Why has economic growth been more pro-poor in some states of india than others? J. Dev. Econ. 68, 381–400.
Ravallion, M., Wodon, Q., 2000a. Does child labor displace schooling? Evidence on behavioral responses to an enrolment subsidy. Econ. J. 110, 158–176.
Ravallion, M., Wodon, Q., 2000b. Banking on the poor? Branch location and non-farm rural development in Bangladesh. Rev. Dev. Econ. 4 (2), 121–139.
Ravallion, M., Datt, G., van de Walle, D., 1991. Quantifying absolute poverty in the developing world. Rev. Income Wealth 37, 345–361.
Ravallion, M., Chen, S., Sangraula, P., 2007. New evidence on the urbanization of global poverty. Popul. Dev. Rev. 33 (4), 667–702.
Ravallion, M., Chen, S., Sangraula, P., 2009. Dollar a day revisited. World Bank Econ. Rev. 23 (2), 163–184.
Ravallion, M., van de Walle, D., Dutta, P., Murgai, R., 2013. Try Telling People their Rights? On Making India's Largest Antipoverty Program work in India's Poorest State. Department of Economics, Georgetown University, Washington, DC.
Raventós, D., 2007. Basic Income, the Material Conditions of Freedom. Pluto Press, London.
Rawls, J., 1967. Distributive justice. In: Laslett, P., Runciman, W.G. (Eds.), Philosophy, Politics and Society.In: Series III,London.
Rawls, J., 1971. A Theory of Justice. Harvard University Press, Cambridge, MA.
Ray, D., 1998. Development Economics. Princeton University Press, Princeton, New Jersey.
Reinikka, R., Svensson, J., 2005. Fighting corruption to improve schooling, evidence from a newspaper campaign in Uganda. J. Eur. Econ. Assoc. 3 (2–3), 259–267, Papers and Proceedings of the Nineteenth Annual Congress of the European Economic Association.
Rhys-Williams, J., 1943. Something to Look Forward To. MacDonald, London.
Ricardo, D., 1817. Principles of Political Economy and Taxation, Everyman Edition, London, 1911.
Robbins, L., 1935. An Essay on the Nature and Significance of Economic Science, London.
Robinson, S., 1976. A note on the U-hypothesis relating income inequality and economic development. Am. Econ. Rev. 66, 437–440.
Roche, D., 1987. The People of Paris, an Essay in Popular Culture in the Eighteenth Century. University of California Press, Berkeley.
Rodriguez, F., Rodrik, D., 2001. Trade Policy and Economic Growth, A Sceptic's Guide to the Cross-National Evidence. NBER Macroeconomic Annual 2000MIT Press, Cambridge, pp. 261–324.
Rodrik, D., 1994. King Kong meets Godzilla, The World Bank and The East Asian Miracle. In: Fishlow, A., Gwin, C., Haggard, S., Rodrik, D., Wade, R. (Eds.), Miracle or Design? Lessons from the East Asian Experience. Overseas Development Council, Washington, DC.
Rodrik, D., 2004. Industrial Policy for the Twenty-First Century, Kennedy School of Government Working Paper RWP04-047, Harvard University.
Roemer, J., 1996. Theories of Distributive Justice. Harvard University Press, Cambridge.
Roemer, J., 1998. Equality of Opportunity. Harvard University Press, Cambridge, MA.
Roemer, J., 2013. Economic Development as Opportunity Equalization, Policy Research Working Paper 6530, World Bank.
Roodman, D., Morduch, J., 2009. The Impact of Microcredit on the Poor in Bangladesh, Revisiting the Evidence, Working Paper 174, Center for Global Development.
Rothschild, E., 2001. Economic Sentiments, Adam Smith, Condorcet, and the Enlightenment. Harvard University Press, Cambridge, MA.

Rothschild, M., Stiglitz, J., 1976. Equilibrium in competitive insurance markets, an essay on the economics of imperfect information. Quart. J. Econ. 90 (4), 629–650.
Rousseau, J.-J., 1754. Discourse on the Origin of Inequality, a Discourse on a Subject Proposed by the Academy of Dijon, What is the Origin of Inequality Among Men, and is it Authorised by Natural Law. Translated by G. D. H. Cole.
Rowntree, B.S., 1902. Poverty, a Study of Town Life. Macmillan, London.
Runciman, W.G., 1966. Relative Deprivation and Social Justice, a Study of Attitudes to Social Inequality in Twentieth Century England. Routledge and Kegan Paul.
Sachs, J., 2005a. Investing in Development, A Practical Plan to Achieve the Millennium Development Goals. United Nations Millennium Project, vol. 1, New York.
Sachs, J., 2005b. The End of Poverty. Economic Possibilities for Our Time. Penguin Books, New York.
Sadoulet, E., de Janvry, A., 1995. Quantitative Development Policy Analysis. The Johns Hopkins University Press, Baltimore.
Saez, E., 2006. Redistribution toward low incomes in rich countries. In: Banerjee, A., Benabou, R., Mookherjee, D. (Eds.), Understanding Poverty. Oxford University Press, Oxford.
Saint-Paul, G., 2011. The Tyranny of Utility. Behavioral Social Science and the Rise of Paternalism. Princeton University Press, Princeton.
Sala-I-Martin, X., Doppelhofer, G., Miller, R., 2004. Determinants of long-term growth, a Bayesian averaging of classical estimates (BACE) approach. Am. Econ. Rev. 94 (4), 813–836.
Sandmo, A., 2015. Income distribution in the history of economic thought. Handbook of Income Distribution, vol. 2A (Atkinson, A.B., Bourguignon, F. Eds.). Elsevier Science, Amsterdam (Chapter 1).
Sen, A.K., 1970a. Collective Choice and Social Welfare. Holden Day, San Francisco.
Sen, A.K., 1970b. Interpersonal aggregation and partial comparability. Econometrica 38 (3), 393–409.
Sen, A.K., 1973. On Economic Inequality. Clarendon Press, Oxford.
Sen, A.K., 1976. Poverty, an ordinal approach to measurement. Econometrica 46, 437–446.
Sen, A.K., 1980. Equality of what? In: McMurrin, S. (Ed.), Tanner Lectures on Human Values. Cambridge University Press, Cambridge.
Sen, A.K., 1981a. Poverty and Famines. Oxford University Press, Oxford.
Sen, A.K., 1981b. Public action and the quality of life in developing countries. Oxf. Bull. Econ. Stat. 43 (4), 287–319.
Sen, A.K., 1983. Poor, relatively speaking. Oxf. Econ. Papers 35 (2), 153–169.
Sen, A.K., 1985. Commodities and Capabilities, North-Holland, Amsterdam.
Sen, A.K., 1999. Development as Freedom. Alfred Knopf, New York.
Sen, A.K., 2000. Social Justice and the distribution of income. In: Atkinson, A.B., Bourguignon, F. (Eds.), Handbook of Income Distribution. In: vol. 1. Elsevier Science, Amsterdam.
Shapiro, C., Stiglitz, J., 1984. Involuntary unemployment as a worker discipline device. Am. Econ. Rev. 74 (3), 433–444.
Shipler, D., 2005. The Working Poor. Invisible in America. Vintage, New York.
Slesnick, D., 2001. Consumption and Social Welfare. Cambridge University Press, Cambridge.
Small, M.L., Harding, D.J., Lamont, M., 2010. Reconsidering culture and poverty. Ann. Am. Acad. Pol. Soc. Sci. 629, 6–29.
Smith, A., 1759. The Theory of Moral Sentiments. A. Millar, London.
Smith, A., 1776. An Inquiry into the Nature and Causes of the Wealth of Nations. Electronic Classic Edition, Pennsylvania State University, Pittsburgh.
Smith, R.M., 2011. Social security as a development institution? The relative efficacy of poor relief provisions under the english old poor law. In: Bayly, C.A., Woolcock, M., Szreter, S., Rao, V. (Eds.), History, Historians and Development Policy, A Necessary Dialogue. Manchester University Press, Manchester.
Solar, P.M., 1995. Poor relief and english economic development before the industrial revolution. Econ. Hist. Rev. 48, 1–22.
Solow, R., 1956. A contribution to the theory of economic growth. Quart. J. Econ. 70 (1), 65–94.
Sridharan, E., 2004. The growth and sectoral composition of India's middle class, its impact on the politics of economic liberalization. India Rev. 3 (4), 405–428.
Steinberg, Stephen, 2011. Poor Reason, Culture Still Doesn't Explain Poverty, Boston Review January 13.

Stephens, W.B., 1998. Education in Britain 1750–1914. Macmillan, London.
Stiglitz, J.E., 1969. Distribution of income and wealth among individuals. Econometrica 37 (3), 382–397.
Stiglitz, J.E., 1974. Incentives and risk sharing in sharecropping. Rev. Econ. Stud. 41, 219–255.
Stiglitz, J.E., 2012. The Price of Inequality. W.W. Norton & Co., New York.
Strauss, J., Thomas, D., 1998. Health, nutrition and economic development. J. Econ. Lit. 36 (2), 766–817.
Strömberg, D., 2004. Radio's impact on new deal spending. Quart. J. Econ. 119 (1), 189–221.
Subramanian, S., Deaton, A., 1996. The demand for food and calories. J. Polit. Econ. 104 (1), 133–162.
Sundquist, J.L., 1968. Politics and Power, the Eisenhower, Kennedy and Johnson Years. Brookings, Washington, DC.
Swamy, A., 1997. A simple test of the nutrition-based efficiency wage model. J. Dev. Econ. 53, 85–98.
Swartz, J., Braxton, D., Viera, A.J., 2011. Calorie menu labeling on quick-service restaurant menus, an updated systematic review of the literature. Int. J. Behav. Nutr. Phys. Act. 8, 135.
Thane, P., 2000. Old Age in English History. Oxford University Press, Oxford.
Thorner, D., 1967. Social and economic studies of Dr Mann. Econ. Pol. Wkly. 2 (13), 612–645.
Tinbergen, J., 1975. Income Distribution, Analyses and Policies, Amsterdam, North-Holland.
Todd, P., Wolpin, K., 2002. Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility, Assessing the Impact of a School Subsidy Program in Mexico, Penn Institute for Economic Research Working Paper 03–022, Department of Economics, University of Pennsylvania.
Townsend, Joseph, 1786. A Dissertation on the Poor Laws by a Well-Wisher to Mankind, Reprint: University of California Press, Berkeley and Los Angeles, 1971.
Townsend, P., 1979. Poverty in the United Kingdom, a Survey of Household Resources and Standards of Living. Penguin Books, Harmonsworth.
Tucker, R.S., 1975. Real wages of artisans in London, 1729–1935. In: Taylor, A.J. (Ed.), The Standard of Living in the Industrial Revolution. Methuen, London.
United Nations Development Programme, 1990. Human Development Report. Oxford University Press for the UNDP, New York.
Vakil, C.N., Brahmanand, 1956. Planning for an Expanding Economy. Vora and Company, Bombay.
van de Walle, D., 1998. Targeting revisited. World Bank Res. Obs. 13 (2), 231–248.
Van Horn Melton, James, 1988. Absolutism and the Eighteenth Century Origins of Compulsory Schooling in Prussia and Austria. Cambridge University Press, Cambridge.
Van Praag, B., 1968. Individual Welfare Functions and Consumer Behavior, Amsterdam, North-Holland.
Vinovskis, M., 1992. Schooling and poor children in 19th-century America. Am. Behav. Sci. 35 (3), 313–331.
Voitchovsky, S., 2005. Does the profile of income inequality matter for economic growth? J. Econ. Growth 10, 273–296.
Voitchovsky, S., 2009. Inequality and economic growth. In: Salverda, W., Nolan, B., Smeeding, T. (Eds.), The Oxford Handbook of Economic Inequality. Oxford University Press, Oxford.
Walker, T.S., Ryan, J.G., 1990. Village and Household Economies in India's Semi-Arid Tropics. Johns Hopkins University Press, Baltimore.
Watts, H.W., 1968. An economic definition of poverty. In: Moynihan, D.P. (Ed.), On Understanding Poverty. Basic Books, New York.
Webb, R.K., 1974. Modern England from the 18th Century to the Present. Dodd, Mead and Company, New York.
Weiner, M., 1991. The Child and the State in India. Princeton University Press, Princeton.
Wicksell, K., 1901. Lectures on Political Economy. Routledge and Kegan Paul, London, 1934 English Edition.
Williamson, J., 1985. Did British Capitalism Breed Inequality? Routledge, London.
Williamson, J., 1998. Globalization and the labor market, using history to inform policy. In: Aghion, P., Williamson, J. (Eds.), Growth, Inequality and Globalization. Cambridge University Press, Cambridge.
Williamson, J., 2001. Demographic change, economic growth and inequality. In: Birdsall, N., Kelley, A., Sinding, S. (Eds.), Population Matters. Oxford University Press, Oxford.
Wilson, W.J., 1987. The truly disadvantaged, the inner city, the underclass and the truly disadvantaged. Chicago University Press, Chicago.

Wim, V.O., Halman, L., 2000. Blame or fate, individual or social? An international comparison of popular explanations of poverty. Eur. Soc. 2 (1), 1–28.
Winters, L.A., McCulloch, N., McKay, A., 2004. Trade liberalization and poverty, the evidence so far. J. Econ. Lit. 42 (March), 72–115.
Woodham-Smith, C., 1962. The Great Hunger, Ireland 1845–9. Hamilton, London.
World Bank, 1980. World Development Report, Poverty and Human Development. Oxford University Press, New York.
World Bank, 1986. Poverty and Hunger, Issues and Options for Food Security in Developing Countries. World Bank, Washington, DC.
World Bank, 1990. World Development Report, Poverty. Oxford University Press, New York.
World Bank, 1993. The East Asian Miracle, Economic Growth and Public Policy. Oxford University Press, New York.
World Bank, 1997. India, Achievements and Challenges in Reducing Poverty: Report No. 16483-IN. World Bank, Washington, DC.
World Bank, 2000. World Development Report, Attacking Poverty. Oxford University Press, New York.
World Bank, 2004. World Development Report, Making Services work for the Poor. Oxford University Press, New York.
World Bank, 2006. World Development Report, Equity and Development. Oxford University Press, New York.
World Bank, 2014. Conditional Cash Transfers. http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTSOCIALPROTECTION/EXTSAFETYNETSANDTRANSFERS/0,contentMDK:20615138~menuPK:282766~pagePK:148956~piPK:216618~theSitePK:282761,00.html

CHAPTER 23

The Welfare State and Antipoverty Policy in Rich Countries

Ive Marx*,†, Brian Nolan‡, Javier Olivera§
*Herman Deleeck Centre for Social Policy, University of Antwerp, Antwerp, Belgium
†Institute for the Study of Labor, IZA, Bonn, Germany
‡Department of Social Policy and Intervention and Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, UK
§Institute for Research on Socio-Economic Inequality, RU INSIDE, University of Luxembourg, Luxembourg, Luxembourg

Contents

Abstract

The aim of this chapter is to highlight some key aspects of recent economic research on the welfare state and antipoverty policy in rich countries and to explore their implications. We begin with the conceptualization and measurement of poverty before sketching out some core features and approaches to the welfare state and antipoverty policies. We then focus on the central plank of the modern welfare state's efforts to address poverty—namely, social protection, discussing in turn the inactive working-age population, child income support, in-work poverty, and retirement and old-age pensions. After that we discuss social spending other than cash transfers, the labor market, education, training and activation, and, finally, intergenerational transmission, childhood, and neighborhoods. We also discuss the

Handbook of Income Distribution, Volume 2B
ISSN 1574-0056, http://dx.doi.org/10.1016/B978-0-444-59429-7.00024-8

welfare state and antipoverty policy in the context of the economic crisis that began in 2007–2008 and the implications for strategies aimed at combining economic growth and employment with making serious inroads into poverty. We conclude with directions for future research.

Keywords

Poverty, Antipoverty policy, Redistribution

JEL Classification Codes

I3, I38, D63

23.1. SETTING THE SCENE

23.1.1 Introduction

Seen by some as primarily a manifestation of inequality in the distribution of income and wealth and by others as a distinctive phenomenon, poverty continues to represent a core challenge for rich countries and their welfare states. This is reflected in the substantial body of research on poverty in industrialized countries, both country-specific and comparative, which seeks to capture the extent of poverty and how it is changing over time, understand its nature, and assess the effectiveness of policies and strategies aimed at addressing it. Poverty is widely regarded as a key social concern in most rich countries, not only in terms of the quality of life of those affected but also in terms of their wasted potential, as well as the risks to the social fabric and to social cohesion more generally. (Chapter 22 by Martin Ravallion argues that the notion that poverty should and can be eliminated in such countries is a relatively recent development and also discusses in depth the links between poverty and macroeconomic performance.) While the nature of poverty and how best to tackle it remain hotly contested at a political and ideological level, the focus of research has increasingly been on the effectiveness of antipoverty policies and strategies, which the recent economic crisis has served only to reinforce.

The aim of this chapter is to highlight some key aspects of recent economic research on the welfare state and antipoverty policy in rich countries and to explore their implications. A core theme will be that the way poverty is conceptualized and measured has fundamental implications for how antipoverty policy is thought about, designed, and implemented. We therefore begin Section 23.1 with a discussion of conceptualization and measurement and key patterns and trends (see also Jäntti and Danziger, 2000), before sketching out some core features and approaches to the welfare state and antipoverty policies. Section 23.2 focuses on the central plank of the modern welfare state's efforts to address poverty—namely, social protection, discussing the inactive working-age population, child income support, in-work poverty, and retirement and old-age pensions. Section 23.3 looks beyond social protection to discuss social spending on other than cash transfers, the labor market, education, training, and activation. Finally, intergenerational transmission, childhood, and neighborhoods are addressed. Section 23.4 discusses the welfare state and antipoverty policy in the context of the economic crisis that began in 2007–2008 and the implications for strategies aimed at combining economic growth and employment with making serious inroads into poverty. Finally, Section 23.5 highlights directions for future research.

23.1.2 Conceptualizing and Measuring Poverty

The definition of poverty underpinning most recent research in Europe relates to exclusion from the ordinary life of the society due to lack of resources, as spelled out, for example, in the particularly influential formulation by Townsend (1979). This has also been very influential from a policy-making perspective as evidenced by the definition adopted by the European economic communities in the mid-1980s:

> *The poor shall be taken to mean persons, families and groups of persons whose resources (material, cultural and social) are so limited as to exclude them from the minimum acceptable way of life in the Member State in which they live.*

Poverty from this starting point has two core elements: It is about inability to participate, and this inability to participate is attributable to inadequate resources. Most economic research then employs income to distinguish the poor, with a great deal of research and debate on how best to establish an income cutoff for that purpose. There are also substantial theoretical and empirical literatures on concepts such as social exclusion (Kronauer, 1998) and on the "capabilities" approach pioneered by Sen (1980, 1993), which have implications for how one thinks about and measures poverty. Indeed, a concern with "poverty" *per se* may predominantly be seen as an Anglo-Saxon concern, with concepts such as deprivation and social exclusion more often the focus in countries such as France or Germany and with the "level of living" approach to living standards and well-being of central importance in the Nordic countries (and having much in common with Sen's capabilities approach in general orientation, on which see for example Erikson, 1993).

In comparative analysis, the most common approach to deriving income thresholds has been to calculate them as proportions of median income in the country in question, with 50% or 60% of the median being the most widely used metric. The underlying rationale is that those falling more than a certain "distance" below the average or normal income in their society are unlikely to be able to participate fully in that society, and notable examples from a very large literature adopting this approach are Atkinson et al. (1995) and the OECD's recent studies *Growing Unequal?* (Whiteford, 2008) and *Divided We Stand* (OECD, 2011a). Such research, like that on income inequality, was for many years bedeviled by differences in definition and measures in the data available for different countries, but sources such as the Luxembourg Income Study (LIS) micro database, the figures produced by Eurostat from micro data for the EU countries, and the database of aggregate poverty (and inequality) estimates assembled by the OECD have greatly improved this situation. Differences across countries and trends over time in relative income poverty measured in this fashion have played a central role in European research and policy debate. Chapter 8 in this volume by Morelli et al. presents evidence on trends in such measures to which we will return below.

This approach to deriving income thresholds can be contrasted with the approach taken by the United States, where the existence of a long-standing official poverty line has fundamentally influenced how poverty is debated and how research is carried out. That standard goes back to the 1960s when it was originally based on the cost of a nutritionally adequate diet, multiplied by a factor to take account of nonfood spending, but its key feature is that it has subsequently been uprated in line with consumer prices, rather than linked to average income or living standards. To characterize this contrast as between "relative" versus "absolute" notions of poverty would be to oversimplify, because above subsistence-level notions of what constitutes poverty inevitably reflect prevailing norms and expectations. The key issue in making comparisons over time is whether the poverty standard is fixed in terms of purchasing power—that is, "anchored" at a point in time or increases as average living standards rise. As Lampman (1971) put it in a U.S. context, in fighting a "War on Poverty" one may want to monitor how well one is doing in meeting a fixed target rather than redefining the target as income changes. However, over any prolonged period where average living standards are rising, this may lose touch with the everyday understanding of poverty in the society. Thus, an influential expert panel reviewing the U.S. official measure saw poverty in terms of insufficient resources for basic living needs, "defined appropriately for the United States today" (Citro and Michael, 1995).

The fact that the "anchored" measure has continued to be seen as relevant in the United States—for all its well-recognized and analyzed technical limitations—is in itself a reflection of the fact that growth in median real incomes has been modest there. In Europe, the set of poverty and social inclusion indicators adopted by the European Union (EU) since 2001 have supplemented purely relative income poverty thresholds with ones anchored at a point in time some years earlier and uprated in line with prices. The onset of the economic crisis from 2007 to 2008, when median income and relative income thresholds actually fell in some countries, proved a salutary reminder of the value of such anchored thresholds. Similar arguments apply in making comparisons across countries at rather different levels of average income: Neither purely country-specific relative measures nor common thresholds tell the whole story with respect to poverty. In a European context, this was brought to the fore by the accession to the EU in 2004 and 2007 of new eastern countries with much lower levels of average income than the "old" member states.

Alternative ways of establishing an income poverty threshold in a rich country have been proposed, for example, by reference to what it costs to buy a specified basket of goods and services, to ordinary expenditure patterns, to standards implicit in social security support rates, or to views in the population about, for example, the income needed to "get by." This continues to represent a significant theme in poverty research literature, as shown by recent attempts to apply the "basket of goods" approach in a consistent fashion across a variety of European countries (for a discussion of strengths and limitations of these alternatives see Nolan and Whelan, 1996). However, the extent to which this research has affected policy formulation and debate remains quite limited, with the relative and anchored income lines dominating. One suspects this is because of their reasonably straightforward empirical derivation.

In a similar vein, the way household size and composition are taken into account in applying those income lines is, for the most part, rather straightforward. The household is conventionally taken as the income recipient unit, as in the study of income inequality more broadly, assuming that income is shared so members reach a common standard of living. The fact that the types of household identified as poor (much more than the overall poverty rate) can be highly sensitive to the precise equivalence scale employed has been known for some time (Buhmann et al., 1987; Coulter et al., 1992), but in the absence of a more satisfactory alternative emerging from research practice, one has to rely on several commonly used scales (the square root of household size, the "OECD scale," and the "modified OECD scale") and (at best) present results with more than one so that this sensitivity can be assessed. While a number of studies have sought to open up the household "black box" from a poverty perspective, a subset of the research on intra-household inequality, more broadly discussed in Chapter 16 of this volume by Chiappori and Meghir, has had little impact on practice in empirical analysis and policy formulation.

The same could be said of the extensive literature on how best to capture the extent of poverty in a single summary indicator, where despite the considerable literature developing sophisticated indicators the most commonly used measure remains the simple headcount. As long ago as in the mid-1970s, Amartya Sen highlighted how the policy maker is faced with the perverse incentive to target the least poor, and Sen's (1976) and alternative ways of incorporating the "poverty gap" and inequality among the poor have been debated, often derived from a set of axioms representing *a priori* notions of the properties such a measure should have. The Foster et al. (1984) class of poverty measures, for example, are additively decomposable and, additionally, allow for different judgments regarding the importance attached to the extent on inequality among the poor. Such poverty measures that capture poverty intensity also suffer from greater sensitivity to measurement error, especially in the presence of extreme low incomes, which often reflects misreporting,[1] and as Myles (2000) argues, their mathematical representation may have made their meaning obscure to potential users. The robustness of poverty

[1] The poverty gap measure advanced by Hills (2002), based on the distance between the threshold and the median income of the poor, is one response to that problem.

orderings has also been a long-standing concern in the literature (Atkinson, 1987; Zheng, 2000), and dominance approaches developed for income inequality comparisons have been adapted for use in the poverty context (see Duclos and Makdissi, 2005), but once again this has not entered mainstream empirical practice, where the comparison of poverty headcounts over time or across countries on the basis of one or, at most, a very limited set of thresholds and equivalence scales remains the norm. An awareness of the importance of measurement error and the need to take statistical confidence intervals seriously in such comparisons does appear to be increasing, however (see Goedemé, 2013). There have also been significant improvements in the quality and comparability of income data for poverty analysis in recent years (as is the case for the analysis of income inequality more generally, as brought out in Morelli et al.'s Chapter 8 and in Tóth, 2014), largely due to the efforts of organizations such as the OECD, the LIS, and Eurostat as well as national statistics offices.

A substantial strand in recent research on poverty that is increasingly influencing practice has focused instead on questioning what economic research had tended to take for granted: that current income is the most satisfactory, or least bad, yardstick available for identifying the poor. It has instead been argued forcefully that low-income fails in practice to distinguish those experiencing poverty and exclusion, because current income does not capture the impact of savings, debt, previous spending on consumer durables, owner-occupied housing, goods and services provided by the State, work-related expenses such as transport and child care, and geographical variation in prices, because needs also differ in ways missed by conventional equivalence scales (for example in relation to disability), and because income from self-employment, home production, and capital are particularly difficult to measure accurately. One response is to measure financial poverty in terms of consumption rather than income, on the basis that the transitory component is a great deal smaller, but expenditure as measured in household budget surveys often covers only a short period and is not the same as consumption, while low expenditure may be associated with saving and does not necessarily capture constrained resources. Other avenues explored in research have been to impute income from durables, owner-occupied housing and noncash benefits, to broaden the needs incorporated into equivalence scales and to combine survey and other data to improve the measurement of income.

The exploitation of longitudinal data has also been a significant contributor to income-based poverty research. Poverty measures are often based on the income of the household in a specific week, month, or year, but (even if measured accurately) income at a particular point in time may not be representative of the usual or longer term income of the household. Longitudinal data tracking households and their incomes have now become much more widely available, allowing those who move in and out of low income to be distinguished from those who are persistently of low income, and a dynamic perspective on income now plays a central role in research on poverty. Bane and Ellwood

(1986) pioneered research on the length of spells in poverty in the United States, and cross-country analysis was pioneered by Duncan et al. (1993). Comparative studies of income poverty dynamics since then include OECD (2001), Whelan et al. (2003), Fouarge and Layte (2005), and Valletta (2006). Movements in and out of poverty are special cases of more general income mobility, discussed in Chapter 10 by Jäntti and Jenkins in this volume. Available studies show what the OECD (2001) has summarized as the seeming paradox that poverty is simultaneously fluid and characterized by long-term traps. Many spells in poverty are short and represent only transitory setbacks, and considerably fewer people are continually poor for an extended period of time than are observed in poverty at a point in time, but on the other hand, the typical year spent in poverty is lived by someone who experiences multiple years of poverty; comparison across countries has found poverty persistence to be particularly high in the United States and much lower in countries with lower cross-sectional poverty rates. The EU's social inclusion indicators now include a measure of persistent poverty, the percentage below the relative poverty threshold in the current year and at least two of the three previous years. More generally, this aspect of poverty research, with its emphasis on trying to understand not only once-off poverty entries and escapes but also the cumulative experience of poverty over years, has had a major impact on the way policy effectiveness is thought about and assessed.

As well as broadening the measurement of income/financial resources and their dynamics, a parallel development in recent poverty research has sought to go beyond income, with a view to:

- identifying the poor more accurately and understanding the causal processes at work,
- capturing the multidimensional nature of poverty, and/or
- encompassing social exclusion conceived as something broader than "financial poverty."

Nonmonetary indicators of deprivation have been used for quite some time to directly capture different aspects of living standards and social exclusion (either on their own or combined with low income), to validate an income poverty threshold, and/or to bring out graphically what it means to be poor; the review of the literature on measures of material deprivation in OECD countries by Boarini and Mira d'Ercole (2006) listed more than 100 studies. Over the past decade or more, nonmonetary indicators measured at micro levels are also increasingly being used to capture the multidimensional nature of poverty and of social exclusion more broadly—especially in Europe, where the concepts of social exclusion and social inclusion have come to be widely used alongside poverty in research and policy circles, unlike in the United States where they have so far had little purchase. Comparative analysis of datasets such as the European Community Household Panel Survey (ECHP) organized by Eurostat and carried out in most of the (then) EU-member states from the mid-1990s to 2001, and the EU-Statistics on Income and Living Conditions (EU-SILC) data-gathering framework, which replaced it, has identified distinct dimensions of disadvantage (see Eurostat, 2005; Guio, 2009; Guio and Macquet, 2007; Nolan and Whelan, 2010, 2011; Whelan et al., 2001), bringing out that low income alone is not enough to predict who experiences poor housing, neighborhood deprivation, poor health and access to health services, and low education. The measurement of multidimensional poverty and inequality, discussed in Chapter 3 of this volume by Aaberge and Brandolini, raises complex issues not only about the best way to identify and empirically capture particular dimensions, but also about how information about different aspects of deprivation or exclusion is best summarized across those dimensions (see Aaberge and Peluso, 2012; Atkinson, 2003; Bourguignon and Chakravarty, 2003; Tsui, 2002).

The focus on multidimensionality has gone well beyond a purely academic concern to also influence the way poverty reduction targets have been framed, both nationally and at EU level. The national poverty reduction target adopted in Ireland in the 1990s, for example, was framed in terms of the combination of low income and "basic" deprivation, and lively debates about how best to frame targets for child poverty in the United Kingdom have centered on the role of multidimensionality. Since 2001 the EU's social inclusion process has at its core a set of indicators designed to monitor progress and support mutual learning that is explicitly and designedly multidimensional, including but going beyond income-based poverty indicators, including indicators of material deprivation and housing deprivation (see Atkinson et al., 2002; Marlier et al., 2007; Nolan and Whelan, 2011; Chapter 3). Even more strikingly, when in 2010 the EU adopted the *Europe 2020* strategy for jobs and growth, which for the first time included poverty reduction among its high-level targets, the target population for poverty reduction was identified as those:

- below the 60% of national median threshold relative income threshold, and/or;
- above the material deprivation threshold, and/or;
- in a jobless household.

A total of 23% of EU citizens were identified as "at-risk-of-poverty and social exclusion," as this was labeled, significantly more than the 16% below the "headline" 60% of median relative income threshold, and EU leaders pledged to bring at least 20 million of these people out of poverty and exclusion by 2020. While once can readily criticize the logic and implications of this precise combination of elements (on which see Nolan and Whelan, 2011), it represents a powerful illustration of the role that multidimensional measures, and direct measures of material deprivation as a central component, have come to play in framing European antipoverty policy.

The European poverty target evolved from a process of development and adoption of social inclusion indicators at EU level over the previous decade (see Atkinson et al., 2002), which has had a significant influence on data and analyses of poverty and antipoverty policy in Europe, and indeed on the way poverty is thought about and research is framed. This serves as an important example of the broader point that a good deal of research on poverty is carried out or sponsored by bodies—national or international—that have an interest in demonstrating that particular sets of policies or orientations toward antipoverty strategy are or are likely to be successful. In a more subtle way, their perspectives will influence the data and indicators available to researchers, and thus the analyses that can be readily undertaken. There have been enormous advances in the availability of accessible micro data in recent years, which has fundamentally influenced poverty research and helped to "democratize" it, but the influence of national governments and international organizations remains substantial.

Finally, in discussing how poverty research is approached, differences in disciplinary perspectives are also important. For example, researchers from an economics perspective are generally more comfortable with financial indicators of living standards and exclusion, and highlight the role of economic incentives in understanding and tackling poverty, whereas sociologists have often been more open to employing nonmonetary measures and highlight the role of social stratification and social context. Having said that, there has been significant blurring of disciplinary boundaries, and poverty research has become a site for particularly fruitful collaborations between *inter alia* economists, sociologists, social policy analysts, geographers, anthropologists, educationalists, epidemiologists, psychologists, and indeed geneticists and neuroscientists, of which this chapter can only give a flavor, concentrating for the most part on the economics literature.

23.1.3 Key Patterns and Trends

As the previous section highlighted, the most common practice in comparative research on poverty remains the application of relative income poverty thresholds and comparisons of headcounts of the proportions falling below those thresholds in different countries. On that basis, poverty rates for various OECD countries based on the data in the LIS have been compared in, for example, Atkinson et al. (1995) and Fritzell and Ritakallio (2004). The OECD has assembled estimates for many of its member countries at intervals from 1980, which have underpinned its important studies in this area (notably OECD, 2008, 2011a) and annual estimates are also now produced by Eurostat for all the member states of the EU. This, together with national data, provides a substantially improved evidence base for the study of poverty across countries and over time.

Chapter 8 in this volume by Morelli et al. summarizes broad trends in relative income poverty over time, with figures from the LIS suggesting that from the mid-1980s to mid-2000s relative income rates generally rose or stayed stable, with very few examples of significant falls. The OECD's analysis of the estimates of relative income poverty it assembled, as examined in Burniaux et al. (1998), Förster and Pearson (2002), Förster and Mira d'Ercole (2005), *Growing Unequal?* (Whiteford, 2008), and *Divided We Stand* (OECD, 2011a), highlighted that the most common direction of change in those figures was upward. The corresponding data produced by Eurostat covers only (most of) the countries in the EU-15 for the period from the mid-1990s to 2001, based on the ECHP, while the expansion of the Union to 27 member states was accompanied by the development of a new statistical apparatus underpinning these estimates, EU-SILC, from about 2004; this means that trends before 2004 can be assessed only for the "old" member states and, for many of these, with a break in the series in the early 2000s, which affects comparability. Nonetheless, the feature displayed by these figures highlighted by a number of studies is the disappointing progress in bringing relative income poverty rates down despite strong growth in employment in some countries over the decade to the mid-2000s (see, for example, Cantillon, 2011).

It is important to note, however, that there is considerable variability in country experiences and that the stability in the overall poverty rate can mask major underlying shifts for different groups. The OECD's studies, for example, show that the trend in relative income poverty for working-age people in the second half of the 1990s and into the 2000s was generally upward, often reflecting a decline in the poverty-reducing impact of taxes and transfers, but pensioners saw sizeable declines in many countries. So policies operating with respect to one important target group—such as older persons—could be having substantial success in reducing poverty while that is obscured by the impact of changes for other groups. In a similar vein, child poverty—the focus of particular attention from policy makers in recent years—may not necessarily move in the same direction as the overall poverty rate, with the U.K. providing an example where trends in child versus overall poverty have deviated substantially over the past two decades.

The OECD has also usefully documented trends in overall poverty taking a threshold "anchored" at 50% of the median in the mid-1980s and then indexed to price changes. On this measure, all OECD countries achieved significant reductions in "absolute" poverty up to year 2000. In countries like Ireland and Spain, which experienced very rapid income growth, poverty in 1995 measured this way was one-sixth the level of 10 years earlier. The U.S. poverty rate on this basis shows a decline from the mid-1980s up until 2000, though smaller than the average decline of the 15 OECD countries included in the study (Förster and Mira d'Ercole, 2005). In a similar vein, it is striking that some countries where relative income poverty remained quite stable or even rose have seen very marked falls in levels of material deprivation, notably some of the lower-income countries joining the EU from 2004 as the common indicators of material deprivation now also produced by Eurostat serve to demonstrate. The evolution of alternative measures of poverty since the onset of the economic crisis across the OECD from 2007 to 2008 is also of central relevance, as we discuss in detail in the final section of this chapter.

National studies for various countries also shed light on poverty trends and the factors at work, though given differences in methods and approaches, it is more difficult to generalize from them. In the United States, for example, most analyses of long-term poverty trends focus on the official poverty rate, which is not linked to average or median income (see Hoynes et al., 2006; Meyer and Wallace, 2009; Smeeding and Thompson, 2013). This (and variants of it) was higher in the 1980s than in the 1970s but despite subsequent falls was still as high in the mid-2000s as it had been in the mid-1970s. Stagnant median wage growth, rising inequality, and the evolution of unemployment have been highlighted in studies, with the changing wage distribution assigned a central role in explaining poverty trends. Studies of poverty trends in the United Kingdom, by contrast, have generally focused on relative income poverty and have highlighted the role of changes in the transfer and direct tax systems in the increase recorded in the 1980s and into the 1990s and then stabilization from the late 1990s. However, as Dickens and Ellwood (2003) emphasize in a comparative study of Britain and the United States, the factors influencing poverty trends can differ substantially between absolute and relative measures as well as countries, and it is hazardous to generalize.

Trends in poverty over time, overall, and for specific subgroups offer one important window into the causal factors involved and into "what works" in addressing poverty, especially in terms of the impact of changes made in social protection and tax systems. It is also striking that the ranking of countries in terms of relative income poverty rates tend to be fairly stable over time. Table 23.1 shows the percentage of people in households falling below 50% and 60% of median (equivalized) disposable household income in 25 OECD countries around the mid-2000s. The simple fact that there is considerable cross-country variation in poverty measured this way—with some countries displaying percentages below 60% of the median as low as 11–12% and at the other extreme countries having figures twice that high—and that the ranking of countries tends to be reasonably stable over time, suggests that there are important structural factors at work from which antipoverty strategies have much to learn.

A similar point is brought home by reference to the variation across countries in relative income poverty rates for specific population subgroups. Table 23.2 illustrates this with the rates for children and older persons falling below 50% of national median income, compared with the population as a whole. Children have above-average rates in about half the countries shown, with the gap being particularly wide in the United Kingdom and the United States, but in a substantial minority, their rate is below average. The elderly have an above-average rate in most countries, with substantial variation in the size of the gap, and there are some instances where their rate is well below the average. A similar comparison across the EU 27 using data from EU-SILC shows similar patterns. So, this reinforces the notion that there is much to be learned in policy terms from analysis of the situation and treatment of similar groups in different countries.

The same is true of other groups that are generally thought of as vulnerable. For example, the unemployed face a significantly heightened risk of relative income poverty virtually everywhere, but the gap between them and the employed varies widely across countries. Similarly, single parents often face much higher risks of poverty than couples with one or two children, but that gap varies a great deal. As OECD (2005) points out, in

Table 23.1 Income poverty rates in OECD countries, mid-2000s

Country	% below 50% of median income	% below 60% of median income
Australia (2003)	12.3	20.4
Austria (2004)	7.1	13.4
Belgium (2000)	8.1	16.1
Canada (2007)	11.9	18.7
Czech Republic (2004)	5.8	11.4
Denmark (2004)	5.6	13.2
Estonia (2004)	12.8	20.4
Finland (2004)	6.6	13.7
France (2005)	8.5	14.9
Germany (2007)	8.4	14.6
Greece (2004)	11.9	19.6
Hungary (2005)	7.4	12.5
Ireland (2004)	13.2	22.0
Italy (2008)	11.9	19.7
Luxembourg (2004)	8.9	13.8
Mexico (2004)	18.3	25.5
Netherlands (2004)	6.3	11.8
Norway (2004)	7.1	12.8
Poland (2004)	10.7	17.2
Slovenia (2004)	7.1	11.7
Spain (2007)	13.7	20.3
Sweden (2005)	5.6	12.0
Switzerland (2004)	8.0	14.8
UK (2004)	11.2	19.0
USA (2007)	17.7	24.4

Source: LIS downloaded.

many countries it is not living in single-parent households *per se* that increases risk, but rather the likelihood that the parent is not at work. As we shall see, this type of comparative analysis plays a central role in research aimed at informing antipoverty policies and strategies.

It is also worth noting that although relative income poverty measures are sometimes dismissed as really only capturing inequality, in fact a country (or group within it) can have zero poverty despite substantial inequality. To give concrete examples, in both the Netherlands and New Zealand the incidence of relative poverty among the elderly (with the 50% of median threshold) is close to zero, although there is substantial income inequality among their elderly populations. The redistributive effort required to truncate the distribution at a widely used poverty threshold like 50% of median equivalent income is in fact a fraction of the actual redistributive flows that take place in most countries. In practice, as Figure 23.1 shows, broadly speaking, where inequality in disposable income is

Table 23.2 Income poverty rates for children and elderly in OECD countries, mid-2000s

	% of below 50% of median income		
Country	**Children**	**Elderly (65+)**	**All**
Australia (2003)	14.0	22.3	12.3
Austria (2004)	6.8	9.4	7.1
Belgium (2000)	7.2	15.4	8.1
Canada (2007)	15.0	8.3	11.9
Czech Republic (2004)	10.2	2.1	5.8
Denmark (2004)	3.9	8.5	5.6
Estonia (2004)	15.4	13.5	12.8
Finland (2004)	4.1	10.3	6.6
France (2005)	10.2	7.4	8.5
Germany (2007)	9.3	9.0	8.4
Greece (2004)	12.4	18.8	11.9
Hungary (2005)	9.9	4.0	7.4
Ireland (2004)	15.9	23.8	13.2
Italy (2008)	17.1	11.0	11.9
Luxembourg (2004)	13.5	4.7	8.9
Mexico (2004)	22.2	27.1	18.3
Netherlands (2004)	9.2	2.4	6.3
Norway (2004)	5.3	8.5	7.1
Poland (2004)	15.6	3.5	10.7
Slovenia (2004)	5.5	16.4	7.1
Spain (2007)	17.3	20.7	13.7
Sweden (2005)	4.7	6.6	5.6
Switzerland (2004)	9.3	15.1	8.0
UK (2004)	13.0	16.3	11.2
USA (2004)	22.0	24.2	17.7

Source: LIS downloaded.

high relative income poverty rates tend to be high as well, but similar inequality levels can be associated with quite different levels of relative income poverty.

23.1.4 The Welfare State and Poverty

As Barr (2001) put it, the welfare state combines the role of piggy bank and Robin Hood, providing collective insurance against social risks while also aiming to ameliorate need and poverty. Redistribution can be horizontal, across the life cycle, or vertical between higher and lower incomes. Poverty reduction is by no means the sole criterion against which the success of welfare state institutions would or should be judged—whether at a point in time or over the life cycle—but it would be widely accepted as among the core aims. Research aimed at assessing success or failure in those terms can focus at the aggregate level, at specific population subgroups, or at particular institutional

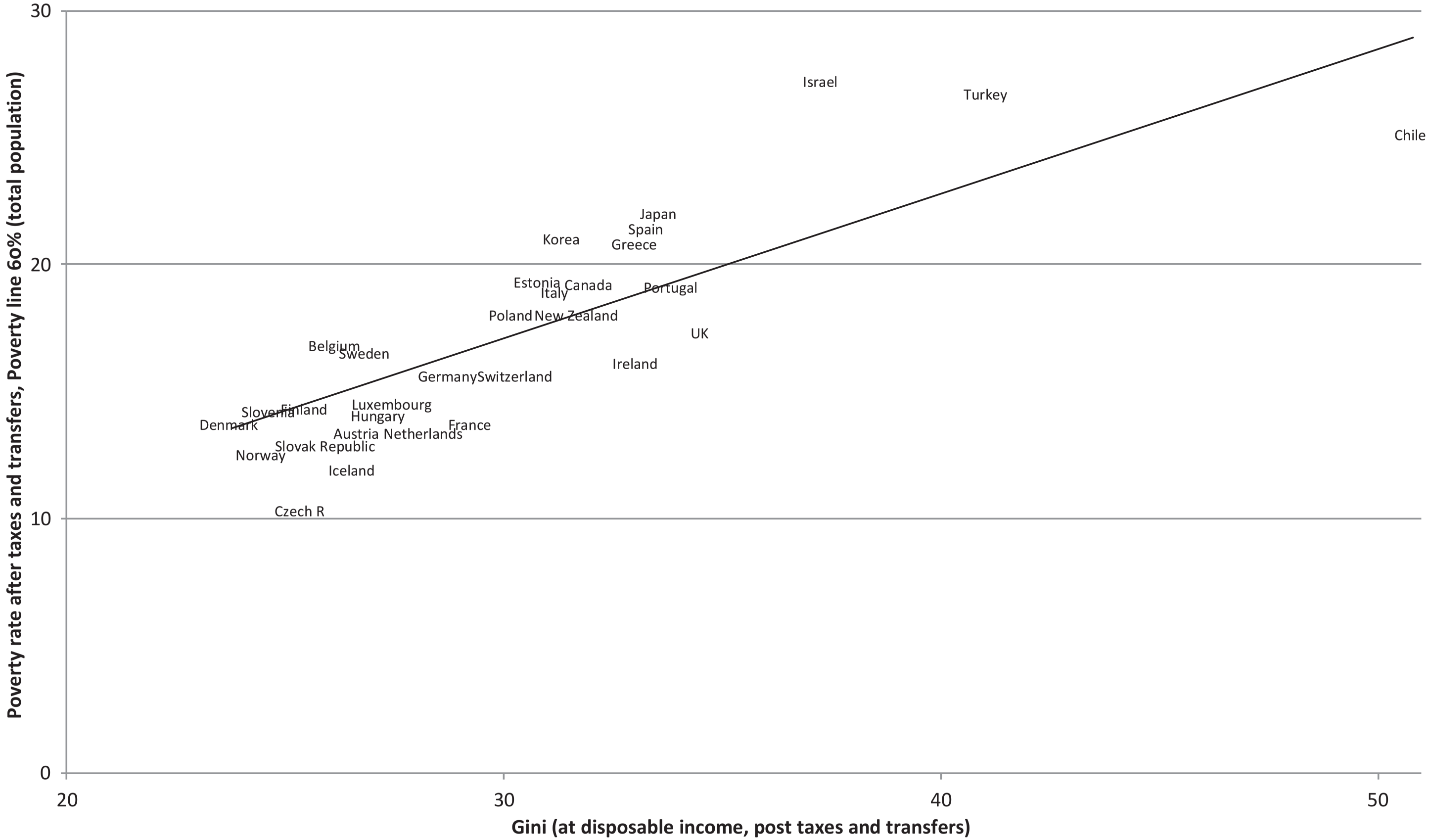


Figure 23.1 Gini coefficient for disposable income and relative income poverty (60% median), 2009, OECD. *Source: OECD income distribution database.*

structures, interventions, or innovations and can be for a particular country or from a comparative perspective.

The nature of that research is also multifaceted. At one end of the spectrum one can locate studies of the effectiveness of very particular aspects of institutional structures or changes in those structures on the target population to whom they are directed. Such evaluation studies employ a wide variety of analytical and technical approaches, which have been the subject of intensive development in the economics literature in recent years. While the outcome studied is occasionally whether people are lifted out of poverty, there is a much more extensive literature focusing on effectiveness in getting unemployed persons into employment, improving performance in school, keeping people out of jail or improving their health, all of which may be expected to impact on poverty status. While randomized controlled trials are recently in vogue in this context—though the negative income tax experiments conducted in the United States and Canada in the 1970s provide early large-scale examples[2]—more commonly, assessments are not based on such an approach. The methods employed include reduced form or limited information models (including least squares, matching methods including propensity score matching, instrumental variable analysis or the closely related regression discontinuity design approach, and difference in difference estimation) versus the estimation of structural models/parameters.[3] Such methods are discussed extensively in other books in this series (notably those focused on labor economics, since assessing the impact of labor market programs has been a particularly fertile field of application); purely from the point of view of research on poverty, though, while influencing specific national reform efforts they have had much less impact on the way antipoverty policy is thought about more broadly.

In that respect, comparative analysis of poverty outcomes and redistributive effort across countries over time continues to dominate (see Cantillon et al., 2014). This is underpinned by the fact that the direct effect of transfers and direct taxes on measured poverty is seen to differ very substantially across countries. OECD analysis concludes that the best-performing countries succeed in lifting about two-thirds of their pre-tax/transfer poor above the threshold, while others only manage to move one-quarter above. Recent EU statistics tell a similar story, as Table 23.3 illustrates: Welfare systems reduce the risk of poverty by 38% on average across the EU, but this impact varies from under 15% to more than 60% across the member states. Some countries achieve better "efficiency" (i.e., reduce poverty more for each euro or dollar spent) through targeting low-income groups, and the role of means-testing is one of the most hotly debated aspects of antipoverty policy to which we return below. However, the prior point to be made here is that

[2] See for example Levine et al. (2005).

[3] For discussion of the advantages and disadvantages of alternative approaches see Chetty (2009), Deaton (2010), Heckman and Urzúa (2010), Imbens (2010), and Heckman (2010).

Table 23.3 Income poverty rates pre- and post-transfers in EU countries, 2007

	Pre-transfer poverty	Post-transfer poverty	Reduction in poverty	
Country	**%**	**%**	**% point**	**%**
Belgium	27.5	15.2	12.3	44.7
Bulgaria	25.5	22	3.5	13.7
Czech Republic	20.1	9.6	10.5	52.2
Denmark	27.1	11.7	15.4	56.8
Germany	24.8	15.2	9.6	38.7
Estonia	25.2	19.4	5.8	23
Ireland	33.1	17.2	15.9	48
Greece	23.7	20.3	3.4	14.3
Spain	23.9	19.7	4.2	17.6
France	26.4	13.1	13.3	50.4
Italy	24.1	19.8	4.3	17.8
Cyprus	21	15.5	5.5	26.2
Latvia	27.2	21.2	6	22.1
Lithuania	25.5	19.1	6.4	25.1
Luxembourg	23.4	13.5	9.9	42.3
Hungary	29.3	12.3	17	58
Malta	21.2	14.8	6.4	30.2
Netherlands	20.6	10.2	10.4	50.5
Austria	24.7	12	12.7	51.4
Poland	26.5	17.3	9.2	34.7
Portugal	24.2	18.1	6.1	25.2
Romania	30.9	24.8	6.1	19.7
Slovenia	23.1	11.5	11.6	50.2
Slovakia	18.2	10.6	7.6	41.8
Finland	28.9	13	15.9	55
Sweden	27.5	10.5	17	61.8
UK	29.7	18.6	11.1	37.4

Source: Eurostat downloaded.

the pattern of incomes from the market, taken as the baseline for comparison, will itself be very much influenced by social transfers and indeed by welfare state institutions more broadly. The existence of social transfers allows substantial numbers of households to have no income from the market, which would not be sustainable otherwise, and the welfare state also affects incentives to work and save in many other ways: the "no welfare state" counterfactual is not known.

A favored mode of analysis in comparative studies is to take a set of countries—at a point in time or pooling cross-sections over time—and assess the relationship between poverty outcomes and a wide set of independent variables reflecting population structures, welfare spending levels and aspects of labor market and welfare state institutions. (These parallel, and sometimes overlap, similar studies employing income inequality as dependent variable reviewed in depth in Chapter 19 of the current volume by Förster and Tóth.) Particularly, influential studies in this vein include Korpi and Palme (1998), Moller et al. (2003), and Kenworthy (2011). In such comparative analysis, countries may be taken as individual units of observation, or they may be grouped together into different "welfare regimes," designed to capture key commonalities/differences in welfare state institutions. Esping-Andersen's (1990) distinction of three distinct regimes has been highly influential: the liberal/Anglo-Saxon countries with minimal public intervention and a preference for targeting and reliance on the market, the social democratic/Nordic countries with comprehensive social entitlements, and the continental welfare states with conservative origins built around social insurance but often along narrowly defined occupational distinctions and a significant degree of reliance on the family (see also Esping-Andersen, 1999, 2009). A fourth "southern" regime is also generally distinguished (Ferrera, 1996), and the treatment of the formerly communist countries of eastern Europe is also a matter for debate. The relationship between aggregate social spending and poverty levels looks systematically different for the countries that joined the EU in 2004 versus the "old" 15 members (see Tsakloglou and Papadopoulos, 2002), but treating them as a single "regime" may not be satisfactory. Many empirical studies have brought out the extent to which conventional indicators of (relative income) poverty vary systematically across welfare regimes (for a recent example see Whelan and Maitre, 2010), and highlight the consistently low rates found in Nordic countries compared with the generally high (though varying) ones seen in the liberal and southern European countries. Looking in some detail at the make-up of household income by source, Maitre et al. (2012) show that countries in the Anglo-Saxon/liberal regime were distinctive in the extent to which low-income households were dependent on social transfers, and also in the extent to which that dependence served as a predictor of material deprivation. The social democratic and corporatist regimes were characterized by a more modest degree of welfare dependence among low-income households, while in the southern Mediterranean countries welfare was not strongly associated with low income and was a particularly poor predictor of deprivation.

Aggregate-level comparative analysis of this type suggests that while transfer and tax systems are undoubtedly key in underpinning variations in poverty levels, other institutional features also contribute in the best performers, notably high levels of minimum wage protection and strong collective bargaining compressing wages, more extensive public and subsidized employment as well active labor market programs, higher levels of public spending on education, and so forth (see also Chapter 19 in this book). Disentangling the effect of these various factors is inherently fraught with difficulties, and that is where simulation via tax-benefit models, discussed in detail in Chapter 24 of this volume by Figari et al. may be particularly helpful. The Euromod research program in particular has enabled comparative tax-benefit simulation analysis across the EU (Figari and Sutherland, 2013; Immervoll et al., 2006) with major implications for policy. To take just one example, Cantillon et al. (2003) showed that simply increasing spending on transfers would have a limited impact on poverty in some EU countries because much of it would go to those already above the poverty line, particularly in the southern European welfare states where pensions dominate.

Another central strand of comparative poverty research has focused on analysis of the characteristics associated with being in poverty and the underlying processes involved, employing micro data. This has been the subject of a very wide variety of studies covering many countries, both descriptive and econometric. Broadly speaking, the types of individual or household seen as at particular risk of poverty include those with low levels of education and skills, the low paid, the unemployed, people with disabilities, single parents, large families, the elderly, children, ethnic minorities, migrants, and refugees. However, there is substantial variation across countries in the patterning of risk, with major implications for how the underlying processes are understood and for policy. The extent to which individual characteristics, qualifications, or experiences manifest themselves in high-poverty rates is clearly seen to depend on the household, labor market, and institutional settings in which those "disadvantages" are experienced. To take one example, the poverty risk for the unemployed compared with others is seen to depend on whether they have dependants, whether there are others in the household at work, and how the welfare state and its institutions try to cushion the impact of unemployment, most importantly through social protection. Strikingly, a high employment rate is clearly not a sufficient condition for low poverty among the working-aged population, which as we discuss below is of central relevance when boosting labor market participation is at the heart of antipoverty policy in many countries.

Finally, the availability of longitudinal data has also allowed the development of econometric modeling of poverty dynamics, which seeks to link observed movements into or out of poverty over time to changes in the earnings, labor force participation, and composition of the household. Duncan et al. (1993) were the first to do so in a comparative setting. A distinction is often made in such dynamic analyses between income "events," such as changes in earnings or benefits, and demographic "events," such as the arrival of a new child, partnership formation, death, marital dissolution, or offspring leaving home. The comparative dynamic analysis by OECD (2005) suggests that changes in household structure may be less important in poverty entries and escapes in European countries than in the United States, with changes in transfers as well as earnings seen to be important in the EU and to a lesser extent in Canada, but much less so in the United States.

23.2. SOCIAL PROTECTION AND REDISTRIBUTION

23.2.1 Introduction

Cash spending as a percentage of GDP is the most widely used measure of how much "effort" is being made to directly redistribute income. Despite its widespread use, this measure has some well-documented shortcomings. First, it ignores the need to jointly analyze benefit and tax policies. Conventional measures of (gross) social expenditure tend to overestimate the cost of welfare in Denmark, Finland, and Sweden, where a substantial amount of benefit spending is clawed back through taxation. Conversely, in the Czech Republic and Slovenia, a substantial share of social spending takes the form of tax breaks for social purposes rather than cash transfers (Adema et al., 2011). Another widely acknowledged weakness of this measure is that it is a very imperfect indicator of policy intent and policy design. A high level of spending may result from very generous benefits flowing to small numbers of people and not necessarily people occupying the bottom end of the distribution—for example, government elites. Yet it may also result from relatively small benefits flowing to a large number of people (De Deken and Kittel, 2007).

Yet, several studies have established a strong empirical relationship at country level between the overall level of social spending and various measures of inequality and inequality reduction, including (relative) poverty. This is arguably one of the more robust findings of comparative poverty research over the past decades (Atkinson et al., 1995; Ferrarini and Nelson, 2003; Gottschalk and Smeeding, 1997; Immervoll and Richardson, 2011; Kenworthy, 2004, 2008, 2011; Kraus, 2004; Nolan and Marx, 2009; OECD, 2008; Pestieau, 2006). Notable in these analyses is that no advanced economy achieved a low level of inequality and/or relative income poverty with a low level of social spending, regardless of how well that country performed on other dimensions that matter for poverty—namely, employment. Contrarily, countries with relatively high social spending tended to have lower inequality and poverty. Here the extent of cross-country variation is always more significant, with some countries achieving more limited inequality/poverty reductions despite high social spending.

The number of countries for which internationally comparative data are available has increased over recent years recently. As Figure 23.2 shows, there are now a number of countries (Czech Republic, Slovakia, Slovenia, as well as Korea) that do combine fairly low levels of social expenditure with low relative poverty rates and income inequality. For the Central European countries, part of the explanation may lie in a reliance on tax breaks as social policy tools, which are not captured in gross social spending indicators. More generally, the redistributive impact of taxes is not captured here (Verbist, 2004; Verbist and Figari, 2014).

This relatively strong relationship between social spending and poverty at the country level probably does not simply reflect the direct impact of transfers only: High-spending countries have other institutional features that contribute, notably high levels of minimum wage protection and strong collective bargaining compressing wages (hence limiting overall inequality), more extensive public and subsidized employment as well active labor market programs, higher levels of public spending on education, and so forth. Disentangling the effect of these various factors is inherently fraught with difficulties. There may in fact be mechanisms of mutual reinforcement between these factors

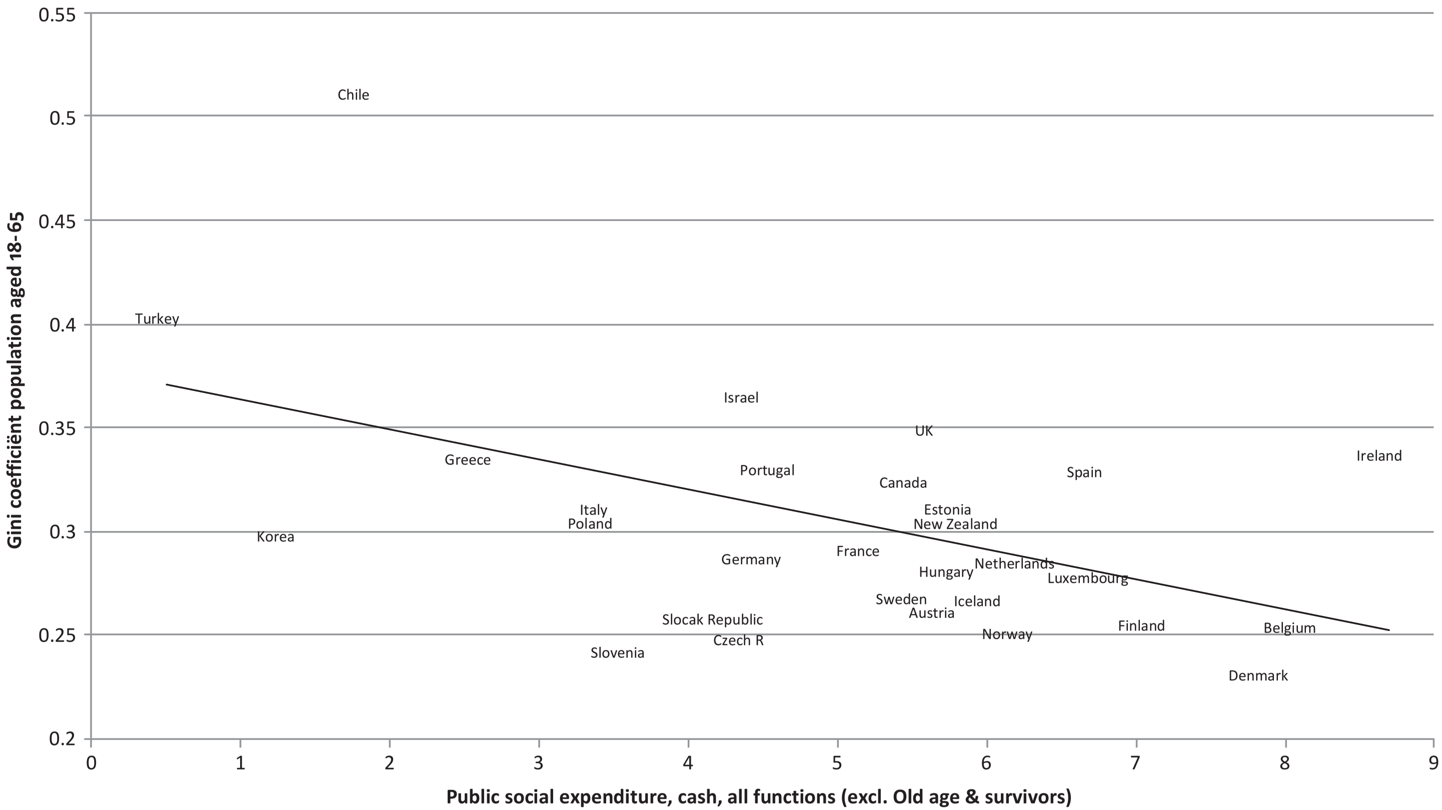


Figure 23.2 Cash public social expenditure and income inequality on the working age in OECD countries, 2009. *Note: Gini coefficient of equivalized disposable household income among the population aged 18–65. Source: OECD Divided We Stand (Gini); OECD SOCX (social expenditure).*

(Beramendi Alvarez, 2001). Barth and Moene (2009) argue that a more equal wage distribution leads to welfare generosity through a process of political competition. In turn, more income redistribution produces more equality. The authors hypothesize that this "equality multiplier" operates mainly through the bottom of the income distribution: The amplification occurs where wages near the bottom of the distribution are compressed, not where higher incomes are compressed. They find empirical support in their analyses on 18 OECD countries over the years 1976–2002.[4]

While in theory, low or moderate levels of social spending could produce low poverty rates if resources were well targeted, the reality remains that almost no advanced economy achieves a low (relative) poverty rate, or a high level of redistribution, with a low level of social spending. Large, universal welfare systems, while on paper being least distributive, distribute in fact the most. Systems that by design strongly target resources to toward the poorest tend to be in fact less redistributive. Korpi and Palme (1998) have called this the "paradox of redistribution."

There is a long-standing controversy in welfare state literature on the question of whether targeting benefits toward the bottom part of the income distribution actually enhances the redistributive impact of welfare state policies, especially of social transfer policies. This issue is of far more than academic importance. In its 2011 *Divided We Stand?*, the OECD states that "redistribution strategies based on government transfers and taxes alone would be neither effective nor financially sustainable." In this context, the OECD (2011a) calls for "well-targeted income support policies." Organizations like the IMF and the World Bank have long advocated targeted benefits. The issue of targeting will probably gain even more poignancy in a post-crisis period marked by continued and, in some cases, increased budget austerity.

The debate on targeting is still marked by opposed views. On the one side there are those who believe that a welfare state can only fight poverty effectively and efficiently (i.e., cost-effectively) when benefits are mainly targeted to those most in need—that is, when benefits are selective. The straightforward argument here is that selective benefit systems are cheaper because fewer resources are "wasted" on people who are not poor. Lower public expenditures imply lower taxes, which in turn are said to be conducive to economic growth. Economic growth, the argument proceeds, benefits the poor directly (although not necessarily proportionally so) and increases at the same time the fiscal base for redistributive policies.

This view of selectivity has never been commonly shared. Two sorts of arguments underpin this more critical stance. First, there are technical considerations. Van Oorschot (2002) sums up the most important dysfunctions of means-testing. First, these

[4] There is a sizeable political economy literature on this issue. McCarty and Pontusson (2009) review a number of political economy theories with regard to voter behavior under different conditions of economic inequality.

include higher administrative costs. Establishing need or other relevant criteria require monitoring, whereas universal benefits allow for less complex eligibility procedures. Furthermore, means-tested benefits are subject to higher non-take-up, partly because of stigmatization issues. Finally, and perhaps most importantly, targeted benefits can give rise to poverty traps, where benefit recipients have little incentive to work because this would entail loss of benefits.

A second line of counterargument is that proponents of selectivity pursue a "mechanical" economic argument that makes abstraction of the political processes, which determine how much is actually available for redistribution. The reasoning is that, paradoxically, in countries with selective welfare systems fewer resources tend to be available for redistribution because there is less widespread and less robust political support for redistribution. As a consequence, the redistributive impact of such systems tends to be smaller. To put it differently, some degree of redistributive "inefficiency" (the Matthew-effect) is said to foster wider and more robust political support for redistribution, including to the most needy. This follows from the fact that a universal welfare state creates a structural coalition of interests between the least well-off and the politically more powerful middle classes (median voter theorem). By contrast, a selective system entails an inherent conflict between the least well-off, by definition the sole recipients of social transfers, and the better-off, who fund the system without the prospect of getting much out of it.

The juxtaposition outlined above forms the starting point for Korpi and Palme's highly influential "Paradox of Redistribution," a paper in which they claim that more selective systems, paradoxically, have a smaller redistributive impact than universal systems offering both minimum income protection as well as income security and cost compensations (for children) in a broader sense. Korpi and Palme (1998) find that, in effect, this relationship is mediated by the relative size of available means for redistribution. Countries with selective redistribution systems, they argue, spend less on redistribution, at least in the public sector. In essence, selective systems are generally smaller systems.

The degree of redistribution is measured here by comparing the actually observed income inequality or at-risk-of-poverty rate with a rather unsophisticated "counterfactual" distribution (Bergh, 2005). In theory this counterfactual ought to accurately reflect the income distribution that would prevail in the absence of social transfers. However, the construction of this counterfactual is hampered by theoretical and practical problems. In most cases, including in Korpi and Palme's paper, pre-transfer income is simply calculated by deducting observed social transfers and re-adding observed taxes. Full abstraction is thus made of any behavioral effects that a change in transfer/tax regime would entail. While patently less than perfect, the reality is that no satisfactory method exists to adequately model such behavioral effects. Many studies have pursued similar empirical approaches—for example, Nelson (2004, 2007).

Another critique has been formulated by Moene and Wallerstein (2003) who have argued that analyses of redistribution need to be done at a more disaggregated level than "the welfare system" because the determining redistributive principles may differ substantially for, say, unemployment, health care, or pensions. Some schemes may rest heavily on the insurance principle, while others may put more weight to the need principle. Universality and selectivity can coexist within one system. Yet, Moene and Wallerstein (2001) also conclude that universal provisions provoke the largest political support because of the higher chance of middle-class citizens to become a beneficiary. Some opinion-based studies also confirm that universal welfare schemes enjoy broader support (Kangas, 1995).

Some recent studies, however, claim that the link between redistribution and universal provision has substantially weakened, or even reversed over time. Kenworthy (2011) reproduces and updates Korpi and Palme's analyses, which related to the situation in 11 countries as of 1985. Kenworthy's findings confirm that countries with more universal benefits achieved more redistribution (measured in the size of redistributive policies in the budget) for the period from 1980 to 1990. By 1995, the image becomes less clear. Data for 2000 and 2005 seem to indicate that there is no longer any association (either positive or negative) between the two variables. Evidently, the findings are based on a small number of cases, which make them particularly sensitive to outliers. A trend toward more targeting in Denmark, in conjunction with an evolution toward more universal benefits in the United States, is largely responsible for the shift in conclusions. Moreover, the new findings may be driven to some extent by the growing share of pensions in social spending. Kenworthy (2011, p. 58) writes about this: "This by no means settles the question, but it does suggest additional reason to rethink the notion that targeting is an impediment to effective redistribution."

Figure 23.3, taken from Marx et al. (2013b), strengthens the finding that the relationship between the extent of targeting and redistributive may have weakened considerably. Here targeting is captured through the concentration index. This is calculated in a similar way as the Gini coefficient. The more negative the concentration coefficient, the more targeted the transfers, whereas the closer the concentration coefficient is to the Gini, the more universal the transfers are distributed. Australia, the United Kingdom, and Denmark have most negative concentration coefficients and can be characterized as strongly pro-poor. Negative concentration coefficients are found in the majority of the countries, pointing to a substantial degree of targeting. Note however that the term "targeting" suggests that outcomes are due to the characteristics of the system, but this need not be the case. Moreover, the outcomes of a system are highly dependent on the characteristics of the underlying population in terms of sociodemographic characteristics, income inequality, composition of income, and so forth. If, for instance, a benefit is designed in such a way that all children are eligible, but all children are situated in the bottom quintile, then this policy measure may appear as targeted in its outcomes, even though its design may

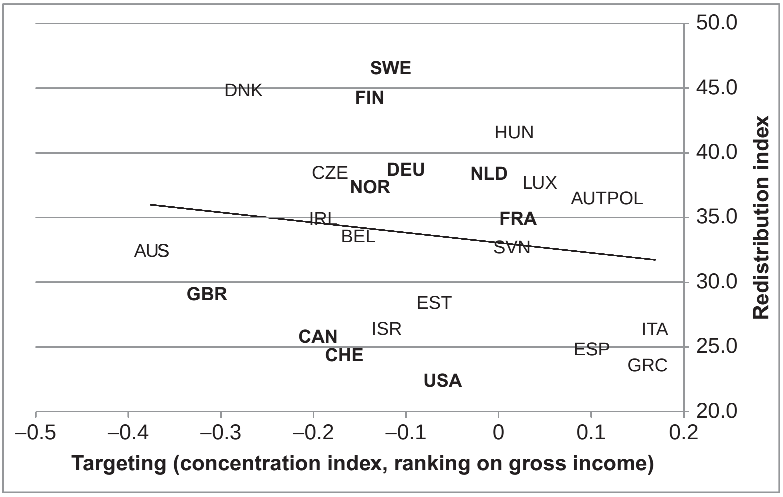


Figure 23.3 Concentration index (ranking by gross income) and redistributive impact, mid-2000s. *Notes: (1) for Belgium, France, Greece, Hungary, Slovenia, and Spain calculations are based on disposable incomes instead of gross incomes due to data availability. (2) The countries included in Korpi and Palme (1998) are in bold. Source: Marx et al. (2013a,b) on the basis of the Luxembourg Income Study.*

not include any means-testing or needs-based characteristics. This means that strictly speaking we cannot derive from the concentration coefficient how pro-poorness of a transfer comes about.

Redistribution refers to the impact of taxes and transfers on income inequality. It is measured by the difference between the Gini coefficients with and without tax-transfers relative to pre-transfer income; this corresponds in this analysis to the difference of the Gini coefficients of market and disposable income relative to that of market income. The impact on inequality is driven by the size of transfers, as well as by their structure, whether these transfers are going relatively more to lower or higher incomes.

Looking more closely at this graph, at the left-hand side are Australia, the United Kingdom, and Denmark, all characterized by having benefit systems that are the most strongly pro-poor of all countries. Yet, the redistributive impact in Denmark appears to be much stronger. Similarly, looking at the countries with still strong pro-poor spending (concentration indices between −0.2 and 0), the corresponding redistributive impact differs a great deal. Some of the countries with the strongest redistributive tax/transfer systems are to be found here (Sweden and Finland), together with some countries with the weakest (the United States, Canada, Israel, and Switzerland). On the right-hand side of the graph—the countries with positive targeting coefficients—the relationship does become consistently negative, especially in the countries with the weakest pro-poor spending (Greece, Spain, and Italy).

Why does a similar degree of strong targeting, as captured by the concentration index, produce stronger redistributive outcomes in Denmark as compared to the United Kingdom and Australia? Similarly, why do similar (quasi) universal systems yield such different redistributive outcomes across countries? This strongly suggests that design features matter. It is notable that one relationship remains fairly strong: the one between the extent of targeting and the size of the system. However, there are exceptions here: A country like Denmark does combine a strong degree of targeting with a high level of social spending.

The strongest redistributive impact is achieved by countries that combine moderate (Sweden and Finland) to strong targeting (Denmark) with comparatively high levels of spending. This suggests that the most redistributive systems are characterized by what is called "targeting within universalism"—that is, systems in which many people receive benefits but where the poorest get relatively more.

It is interesting to note that the very strong relationship between the extent of targeting and the size of the spending has weakened, as is documented by Kenworthy (2011). One of the factors that arguably made targeted systems less politically robust and prone to spending cuts in the 1980s was the fact that strongly targeted (means-tested) benefits entailed strong work disincentives and also (perceived) family formation incentives. The last decades have seen an intensified attention to this issue. To reduce work disincentives, earnings disregards have been introduced for people who make a (partial) transition from complete benefit dependency to part-time work.

Most importantly perhaps, means-tested benefits are no longer exclusively aimed at people not in work, but also at those in work in low-paying jobs. The French RSA (Revenu de Solidarité Active) scheme is a good example of a new style means-tested benefit scheme that offers integrated support for the nonemployed and (part-time) low-paid workers alike. The scheme also has entirely different work incentives. The RSA was introduced in France in 2008 with the specific aim of remodeling the incentive structure social assistance beneficiaries, and particularly to make work or returning to education a more lucrative financial prospect. The previous minimum income system (Minimum Integration Income) was based on a one-for-one trade-off of benefit for earned income. Under RSA, a 62% slope is applied. Efforts have also been made to encourage beneficiaries of RSA into employment, for example, with assisted employment contracts and (improved) insertion mechanisms. In addition, the RSA has simplified the provision of social protection by combining several previously separate schemes into a single sum. A household with no earned income is eligible for the "basic RSA," which is defined at the household level and takes into account the composition of the household. The "in-work RSA" acts as a top-up for people paid less than the national minimum wage (SMIC).

The point here is that targeted, means-tested systems look totally different today from the systems in place in the 1980s. Whereas the old systems were the focus of harsh welfare critiques, especially from the right, the new targeted systems are lauded as gateways of welfare to work. They enjoy broad partisan support, as is evident in the United Kingdom where the working tax credit (WTC), implemented by the Labor government, building on a scheme implemented under a Conservative one, is again expanded by the current Conservative one. Similarly, in France, the newly elected socialist government has no intentions for a major overhaul of the RSA, introduced by the Fillon/Sarkozy government.

In the United States, the earned income tax credit (EITC)—a transfer program for households of low earnings—has become the country's preeminent welfare program (Kenworthy, 2011). The system appears to enjoy far broader and more robust political support than earlier U.S. antipoverty programs. The system also is less strongly targeted than earlier provisions, and it caters to larger sections of the electorate, including the (lower) middle class, which may account for that expansion. However, an equally if not more important factor may well be the fact that the system is perceived to encourage and reward work.

23.2.2 Cash Transfers for the Inactive Working-Age Population

Much comparative poverty research that has sought to link observed variation in income inequality and poverty across countries to policy has relied on government (social) spending statistics as indicators of policy "effort." As we have seen, the relationship across countries between the level of social spending as a percentage of GDP, or some related indicator, and observed inequality or poverty levels is in fact by and large a rather strong one. This is in a way surprising because the level of spending is as much reflective of the number of people receiving benefits than it is of the level and thus potential adequacy of those benefits. Likewise, measured outcomes, for example, pre- versus post-transfer differences in inequality or poverty also depend on a host of factors that are independent or only indirectly influenced by policy: contextual and compositional factors, including labor market conditions (unemployment, employment patterns, and wages), household composition (patterns of cohabitation, marriage, divorce, childbirth, etc.), and policies that influence these dynamics (e.g., ALMPs and child care).

If we want to understand variations in outcomes we need more sophisticated and accurate measures of policy effort and policy design than spending indicators. So-called institutional indicators aim to be directly reflective of policy intent and design. Replacement rates for various branches of social insurance are commonly applied indicators of social protection. They are intended to express the level of benefit generosity within a particular provision, for example unemployment or disability insurance. The OECD has been compiling such time series for a considerable length of time. Academic databases have been compiled by, among others, the Swedish Institute of Social Research (the SCIP database) and the University of Connecticut (Scruggs database).

While such indicators are more directly reflective than spending-based measures of what actually happens at policy levels, they are not without their drawbacks. One is that replacement rates are generally expressed as a proportion of a reference wage. This is problematic for various reasons. With the growth of part-time and temporary employment, it has become increasingly difficult to specify a consistent wage denominator on the basis of available data. More importantly, wages have generally not evolved in line with the standard of living (and thus the relative poverty threshold). In many countries the standard of living has increased thanks to the proliferation of dual income families rather than through real wage growth. The mere fact that benefits follow wages says little about the potential adequacy of benefits in terms of poverty relief. A second important problem is that replacement rates, for example, within the systems of unemployment insurance or invalidity, do not capture the entitlement criteria applied, nor do they adequately express the entitlement periods. Nonetheless, there are strong indications that these are precisely the areas where policy makers have intervened the most. Unemployment benefit entitlement, for example, is now linked more strongly with job-search intensity. A third important issue is that replacement rates are based on a narrow rationale and tend to be calculated on a purely individual basis. For example, unemployment benefits may be combined with (increased) child benefit and other allowances. Additionally, of course, there may be the income of other household members, including its impact on benefit entitlement and vice versa. Also relevant in this context is the role of taxation. In most instances, the level of income protection that people actually receive in various situations is determined by a complex interaction between social security, social assistance, and taxation.

It is nevertheless interesting and relevant to consider trends. OECD time series on net replacement rates for the unemployed provide strong indications of reduced cash support for the unemployed between 1995 and 2005 (Immervoll and Richardson, 2011). Seven of the 10 countries recorded declining NRRs. Finland and Germany saw the biggest reductions in net replacement rates. Changes for the unemployed in most countries tended to be less damaging (or, sometimes, more beneficial) for families with children. The largest relative income drop was generally faced by long-term unemployed job seekers who mostly rely on unemployment assistance or social assistance for income support.

In the remainder of this section, we will focus in somewhat more detail on institutional indicators of minimum income protection because adequate protection against severe financial poverty is arguably the first duty of the welfare state and also because poverty relief is the prime focus of this chapter. Such a focus is further desirable because the design features of tax and benefits systems, and especially the way various programs interact in specific situations, tend to be so complex that they are not accurately and validly captured in a limited number of parameters. Minimum income protection provisions also mark the ground floor of other income maintenance provisions; minimum social insurance levels and minimum wages are almost always above the level of the social safety net. In that sense, indicators of minimum income protection also tell us something about the generosity of other income maintenance provisions (Marx and Nelson, 2013).

We draw on the CSB minimum income protection indicators (MIPI) dataset. In this dataset net income packages are calculated using the so-called model family approach, where the income package of households in various situations (varying by household composition and income levels) in simulated, taking into account all relevant benefits for which such households are eligible and also taking into account taxes. The MIPI database is among the most comprehensive databases available in terms of geographic and longitudinal scope, as well as in terms of the range of household situations and income components. It is worth pointing out that such institutional indicators have their limits too. They are calculated for a limited number of family types and situations. The assumption is that there is full take-up of benefits and that people effectively and immediately receive what they are entitled to. In the case of minimum wages, the assumption is these are fully enforced. However, this is not always the case and this is one reason why the observed relationship between generosity levels, as reflected in these indicators and outcomes, is relatively weak.

Van Mechelen and Marchal (2013) have analyzed patterns and trends in the level of minimum income protection for able-bodied citizens in the European countries. The chief focus is on means-tested benefits providing minimum income protection, usually in the form of social assistance. These general means-tested benefits provide cash benefits for all or almost all people below a specified minimum income level. In some countries separate schemes exist for such groups as newly arrived migrants or the disabled. The empirical analyses use data from the CSB-MIPI and cover social assistance developments in 25 European countries and three U.S. states. The study shows that the minimum income benefit packages for the able-bodied in Europe have become increasingly inadequate in providing income levels sufficient to raise households above the EU at-risk-of poverty rate, defined as 60% of median equivalent income in each country (Figure 23.4). The overall tendency for the 1990s was one of almost uniform erosion of benefit levels, relative to the development of wages. This downward trend in the relative income position of families in receipt of social assistance changes somewhat in the 2000s, when the erosion of the level of benefit packages came to a halt in a number of countries. In a few countries, there is even evidence of a partial reversal of the declining trend, thus somewhat strengthening the income position of able-bodied persons that are in receipt of social assistance benefits. During the crisis period in particular, a small number of countries took extra steps to increase protection levels (Marchal et al., 2014). Despite a number of positive developments, net incomes of minimum income recipients continue to fall well short of the EU's at-risk-of-poverty threshold in all but a few EU countries. The size of the gap between the level of the social safety net and the poverty threshold varies across countries and family types, but it is generally quite substantial.

While the erosion of minimum income protection levels seems to have slowed, the fact remains that Europe's final safety nets offer inadequate protection in all but a handful of countries. This begs the question: Why are social safety nets not more adequate? Let

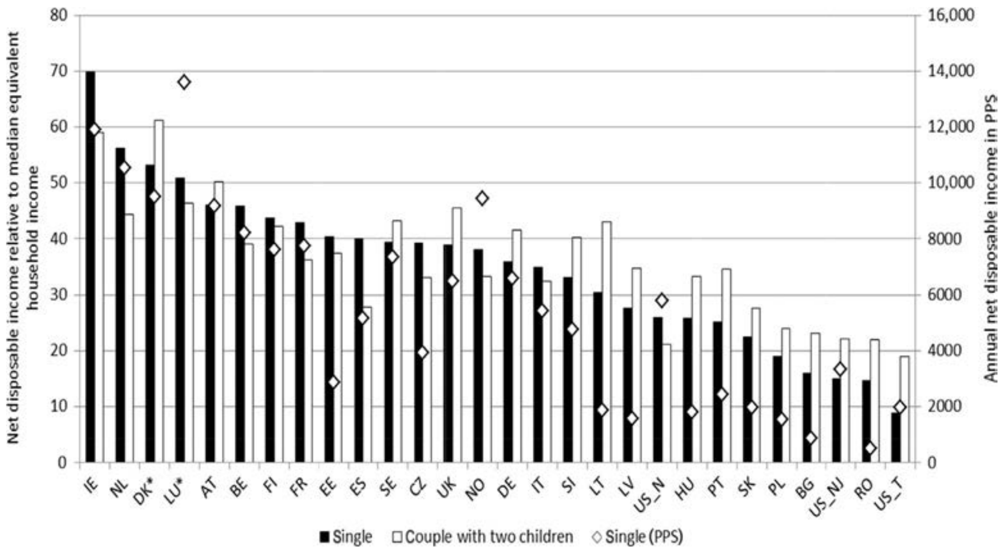


Figure 23.4 The level of the social safety net in the EU and three U.S. states, 2012. *Notes: In some countries, such as the United States, Italy, and Bulgaria, time limits apply, either formal or discretionary. In order to avoid additional assumptions, the levels displayed do not take these time limits into account. Source: CSB-MIPI (Eurostat; US Bureau of the Census and Bureau of Labor Statistics; Van Mechelen et al., 2011).*

us briefly consider two potential impediments: first, "adequate social safety nets are not affordable" and second, "adequate social safety nets undermine the work ethic and people's willingness to work."

Are adequate social safety nets too costly? Final safety net provisions (social assistance schemes) generally constitute only a fraction of total social transfer spending (typically well below 2.5% of GDP in Europe, except in Ireland and the UK), with the bulk of outlays going to pensions, unemployment and disability insurance, child benefits, and other benefits. Vandenbroucke et al. (2013) have made tentative calculations showing that the redistributive effort required to lift all equivalent household incomes to the 60% level would be below 2.5% of aggregate household income in most European countries and nowhere higher than 3.5%. The countries that would have to make such a relatively great effort are all southern and eastern member states. Vandenbroucke et al. (2013) also show that it is not the case that being poor in GDP per capita always implies a great redistributive effort to close the poverty gap. The Czech Republic and Hungary are relatively poor in terms of GDP per capita, but closing the poverty gap would require relatively little effort. On the other hand, Denmark and the United Kingdom have much higher living standards, yet they would have to make a relatively sizeable effort to close the poverty gap. Such a mechanical calculation ignores incentive effects and behavioral change (more poor people may prefer social assistance to low-paying jobs; the nonpoor may reduce their work effort). The real cost of such an operation is probably higher than the mechanical effect and the calculation may be seen as indicating a lower boundary for the distributive effort that is required. Still, the calculation also illustrates that the cost of an adequate social safety net is not necessarily outside of the realm of the conceivable.

Are adequate social safety nets compatible with work incentives? Despite widespread and sometimes strongly worded concerns over the potential work disincentive effects of social safety nets, empirical studies tell a more nuanced story (Immervoll, 2012). The income gap between situation of full-time dependence on minimum income benefits and a full-time job at the minimum wage (or the lowest prevailing wage) is in fact quite substantial in most European countries, especially for single persons. In some countries and under certain circumstances, particular groups such as single parents with young children gain relatively little from moving into a low-paid job, especially when child care costs are accounted for. Partial transitions into work—moving to a small part-time job—also do not pay in certain circumstances. But generally speaking it is hard to argue that long-term dependence on social assistance benefits is an attractive financial proposition in most of Europe. The hypothetical Europe-wide introduction of social assistance minimums equal to 60% of median income would, however, create a financial inactivity trap in many countries, as is also brought out in the paper by Vandenbroucke et al. (2013). In countries such as Bulgaria, Estonia, Slovenia, and Lithuania, the net income of a single benefit recipient would be between 25% and 30% higher than the equivalent income of a single person working at minimum wage; in Spain and the Czech Republic, the relative advantage of the benefit claimant would amount to around 15%. This implies that if such countries would wish to move toward better final safety net provisions then minimum income floors would have to be raised at least in step.

This would require quite substantial increases in minimum wages. In 2013, 20 member states of the EU had a national minimum wage set by government, often in cooperation with or on the advice of the social partners, or by the social partners themselves in a national agreement. As is illustrated in Figure 23.5, presenting figures for 2010, only for single persons and only in a number of countries do net income packages at minimum wage level (taking into account taxes and individual social security contributions, but also social benefits) reach or exceed the EU's at-risk-of-poverty threshold, as in all graphs set at 60% of median equivalent household income in each country. For single parents and sole breadwinners with a partner and children to support, net income packages at minimum wage are below this threshold almost everywhere, usually by a wide margin. This is the case despite shifts over the past decade toward tax relief and additional income support provisions for low-paid workers (Marx et al., 2013a).

When it comes to the question of whether and to what level minimum wages and hence minimum income benefits in general could be increased, opinions clearly diverge. Concerns about the work disincentive effects of social safety nets are legitimate, as are

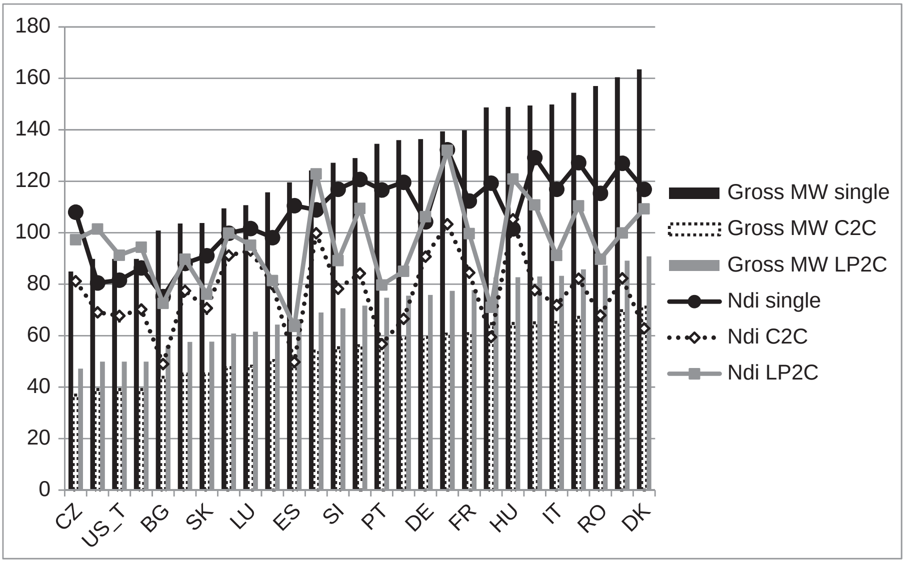


Figure 23.5 Gross minimum wages and net incomes at minimum wage as a percentage of the relative poverty threshold, 2012, selected EU-member states plus the United States (New Jersey). *Source: CSB-MIPI (Eurostat, 2011; US Census Bureau, 2003; Van Mechelen et al., 2011).*

concerns over potential negative employment effects of minimum wages, especially if these were to be set at levels high enough to keep households solely reliant on that wage out of poverty. The fact remains, however, that countries such as Denmark or the Netherlands combine what are comparatively among the highest levels of minimum protection for workers and nonworkers alike with labor market outcomes that on various dimensions are also among the best in the industrialized world. The Netherlands and Denmark enjoy among the highest employment rates in Europe and the lowest (long-term) unemployment rates.

Elaborate active labor market policies, specifically activation efforts directed at social assistance recipients, coupled with intensive monitoring and noncompliance sanctioning, appear to play a key role here. But it appears that the strength of overall labor demand is a key contextual factor for such associated policies and practices to effectively result in low levels of long-term dependence. Moreover, in terms of quality of employment, Denmark and the Netherlands are clearly among the best performers in Europe with relatively few workers in low-quality jobs (European Commission, 2008). Replicating the activation, empowerment, and sanctioning aspects associated with comparatively generous systems may well be difficult enough in itself. Replicating a context where job growth is strong and where jobs are sufficiently rewarding and attractive may be even more difficult.

Relatively elevated social safety nets and other income protection systems can be compatible with well-functioning labor markets. In fact, such systems may actually be conducive to well-functioning labor markets. Flexicurity proponents identify adequate social security benefits as an essential flexicurity pillar in that adequate benefits stimulate and accommodate labor market transitions and reduce risk aversion among workers (Bekker and Wilthagen, 2008).

23.2.3 Child Poverty and Child Cash Transfers

Children are generally at a higher risk of poverty than the population as a whole (Atkinson and Marlier, 2010). In addition, child poverty trends have, for the most part, not been favorable over the past decade (see also Chapter 8 in this book). The latest 2010 EU-SILC data shows that between 2005 and 2010 the at-risk-of-child-poverty rate increased in 17 out of 29 countries (EU27 plus Iceland and Norway). Child poverty rates rose in all the Nordic countries, Germany, and France. In most countries where child poverty fell this was in part the result of a fall in the 60% of median income threshold due the recession (Czech Republic, Estonia, Ireland, Lithuania, Poland, and Portugal). Poverty gaps (the gap between net income and the poverty threshold) for children have also risen between 2005 and 2010 in 15 out of the 29 countries. This deteriorating situation is of course the result of rising unemployment. However, in 2010, the majority of countries in the EU have more than 20% of poor children living in households with all working-age members in employment (work intensity of 1) and all but Belgium, Bulgaria, Czech Republic, Finland, Hungary, Ireland, and the UK have more than half of their poor children living in households with a work intensity of 0.5 or more (Van Mechelen and Bradshaw, 2013).

There are a number of reasons why children are living in poverty when their parents are employed. One explanation is that parental earnings are too low either because they are working part time and/or full-time but their wage is low. The second explanation is that families may be taxed into poverty. The direct taxes taken in income tax and social insurance contributions reduce gross incomes so much that they fall below the poverty threshold. The third explanation is that the cash benefits paid by the state to help parents with the costs of raising children are inadequate. Finally, the reason why a child with a working parent may be poor is that after having paid for housing and other charges the resources available for consumption are too little.

Countries use different mixes of tax benefits and cash benefits for delivering help to families with children. One can distinguish between income-related and universal—that is, non-income-related—cash benefits. Income-related benefits aim to target direct cash transfers to low-income families. Governments may decide to target benefits to other specific groups, for example, single parents or disabled children. Tax instruments are also used to redistribute income from childless families to families with dependent children—either in the form of tax allowances or tax credits specifically aimed at families with children. Tax allowances are deducted from taxable income whereas tax credits are subtracted from the amount of tax due. Tax credits may be wasteable or nonwasteable. Nonwasteable or refundable tax credits are tax benefits that can be paid as cash transfer to the taxpayer whenever the benefit exceeds tax liability. Wasteable tax credits can only be used if tax liability is positive. Both cash and tax benefits tend to vary by the age and the number of children (Bradshaw and Finch, 2002; Van Lancker and Ghysels, 2012).

Child benefit packages, as a whole, play an important role in preventing financial poverty. Nevertheless, in many countries child benefit packages fail to protect low-wage earners against poverty. In all countries the incomes of single-earner couples on minimum wages is below the poverty line. The child benefit package for a lone parent is more generous in most countries. However, how and whether child care costs are subsidized makes a big difference to the package, especially for lone parents. The costs of child care can undermine the value of the package in some countries. Whereas during the 1990s child benefit packages have been able to escape welfare erosion, over the past decade the value of the package relative to median equivalized income has fallen in more countries than it has increased (Van Mechelen and Bradshaw, 2013). This trend of decreasing child benefits has affected both low-paid families and the better-off.

Various studies have looked in detail at the structure of the child benefit package (e.g., Bradshaw, 2010; Bradshaw and Finch, 2002; Corak et al., 2005; Matsaganis et al., 2005; Van Lancker and Ghysels, 2012) and have documented the adequacy of child support arrangements in terms of poverty alleviation using empirical income surveys. Corak et al. (2005) find that universal child-related benefits that also have some degree of targeting at the poorest protect best against poverty. Their conclusion that targeting within universalism yields the best outcomes is echoed by Van Mechelen and Marchal (2013). They find that cross-country variation in the level of child-benefit packages for single-earner families on low pay largely overlaps with the degree of low-income targeting. Model family-type simulations suggest that comparatively generous packages for low-paid workers are to be found in countries where financial help for families with children is well targeted at low-income households by means of income-related cash benefits, refundable income-related tax credits, or social assistance top-ups. However, model family-type simulations effectively assume full take-up of benefits and full granting of rights. In reality, selective benefit systems may be quite ineffective with regard to poverty alleviation due to take-up problems and labor market disincentives (Deacon and Bradshaw, 1983; Gassmann and Notten, 2008). Van Mechelen and Bradshaw (2013) also show that child benefit packages are often also above average in countries with universal cash benefits but are combined with income-related cash benefits, housing allowances, or supplementary benefits from social assistance (Ireland, France, Austria, and Finland). This finding may in effect confirm and reinforce the assertion in empirical literature that targeting may be not so bad, if embedded in a universal social insurance context (Kenworthy, 2011; Skocpol, 1991; Van Lancker and Ghysels, 2012; Whiteford, 2008).

23.2.4 The Working Poor and Combating In-Work Poverty

The issue of in-work poverty has received increased attention recently (Andreß and Lohmann, 2008; Crettaz, 2011; Fraser et al., 2011; Lohmann, 2009; Maitre et al., 2012; Marx and Nolan, 2013; OECD, 2008). It is usually linked to the growth of low-paid insecure employment in the service sector. The contrast is often drawn with the golden years of welfare capitalism when the manufacturing industry provided stable, well-paid employment even for those with little or no formal education. As Esping-Andersen et al. (2002) put it: "We no longer live in a world in which low-skilled workers can support the entire family. The basic requisite for a good life is increasingly strong cognitive skills and professional qualifications . . . Employment remains as always the sine qua non for good life chances, but the requirements for access to quality jobs are rising and are likely to continue to do so." By the same token, Bonoli (2007, p. 496) states, "Postindustrial labour markets are characterized by higher wage inequality with the result that for those at the bottom end of the wage distribution, access to employment is not a guarantee of a poverty-free existence."

At the same time that good jobs for the less skilled are becoming scarcer, an increased policy emphasis on activation has become evident in many European countries, certainly at the level of rhetoric, and gauging by some indicators also in terms of actual policy (Barbier and Ludwig-Mayerhofer, 2004; Digeldey, 2007; Eichhorst et al., 2008; Kenworthy, 2008; OECD, 2007a,b). Within the broad set of activation strategies deployed, an important number specifically target the long-term unemployed, including social assistance recipients. And within this set, an important number of measures are aimed at stimulating these people, who generally have low levels of educational attainment, into relatively low-paid/minimum-wage level jobs.

So has in-work poverty become more prevalent? The literature on the working poor employs a variety of definitions based on different approaches of what is meant by "poor" and by "working" (for an overview see Crettaz, 2011; Nolan and Marx, 2000). The working poor are conventionally defined and measured as those individuals who have been mainly working during the reference year (either in employment or self-employment) and whose household equivalized disposable income is below 60% of the median in the country in question. It is widely recognized that analysis of in-work poverty needs to distinguish between employees and the self-employed, both because of their differing nature and because survey information on self-employment income is normally less reliable than wages and salaries, and also between full-time and part-time workers, which is another important distinction. In fact, with the growth of part-time work, zero-hour contracts, internships, and so forth, "being employed" has become a very fuzzy heterogeneous concept indeed. Moreover, combining two levels of analysis—the individual's labor market status and the household's income (adjusted for household size)—inherently complicates interpretation, because the labor market status of other persons in the household, rather than that of the individual being considered, may be crucial, as may the number of dependent children if any. Using a year as the reference period for labor market status and income position also complicates interpretation. Those working for part but not all of the year may be in poverty on an annual basis for that reason even if they were not poor while working, and how much of the year does one have to work to be counted as "working"? For these and other reasons, this definition/measure makes it difficult to identify the different factors potentially underlying the phenomenon and thus the locus or loci of policy failure, which could include: low (household) work intensity; inadequate out-of-work benefits; inadequate earnings; inadequate earnings supplements, the number of dependent people (children) relative to income, and so on.

Data from the EU-SILC database clearly shows that in-work poverty is a Europe-wide phenomenon. The prevalence of in-work poverty varies across EU countries; the extent of in-work poverty ranges from a low of 4–5% in Austria, Belgium, the Czech Republic, Finland, the Netherlands, and Slovenia up to 13–14% in Greece and Spain and 17% in Romania. On the basis of Eurostat figures, which combine data from ECHP and SILC, we can seek no general tendency for in-work poverty to have risen since the start of the century. Taking the time span from 2000 to 2010, in-work poverty is seen to have increased over the decade in countries such as Denmark, Germany, Spain, Luxembourg, Romania, and Sweden, but fell in as many countries. Abstracting altogether from the crisis period, a comparison of 2000 with 2006 also fails to show a marked rise in in-work poverty in many countries. The common presumption of a rising trend is therefore not supported by this data and indicator. However, the fact that the sources of data for 2000, unlike the later years, are not EU-SILC means that the trends shown have to be treated with some caution.

It is useful to relate these figures and trends to analysis by the OECD, providing a point of comparison and covering the decade from the mid-1990s to the mid-2000s (see OECD, 2009). Drawing on a variety of sources but seeking to apply a uniform methodology, the OECD found in-work poverty to have increased substantially in EU countries such as Germany, the Netherlands, and Luxembourg over this decade, but declined substantially in some other countries such as Italy. The OECD figures also draw on different data sources and employ a different definition—namely, in-work poverty being measured as households below 50% (rather than 60%) of median poverty threshold (with a different equivalence scale), and with "working" being captured at household rather than individual level by the presence of at least one person in work in the household. The study by Airio (2008) of the period 1970–2000 covering six OECD countries (and mostly based on data from the LIS) concludes that it is difficult to find any common trend on in-work poverty. These differences illustrate the care that must be exercised in drawing strong conclusions about levels and trends in in-work poverty across countries, because definitions, data, and period covered can all affect the outcome.

Which policy action, or set of policy actions, is most appropriate cannot be seen as entirely independent from normative notions that underlie the various ways the causes of working-age poverty in relation to work can be construed. Take for example a dual adult household with only one working adult and three dependent children. The male breadwinner has a low-paid job, yet is paid well above the minimum wage. Child benefits are limited. Whether their at-risk-of-financial-poverty status is construed as a problem of insufficient breadwinner earnings, or as a problem of partner nonparticipation, or as a problem of insufficient child support makes a fundamental difference as to what type of policy action is to be examined and possibly favored. In the case of traditional breadwinner-type households with insufficient earnings, the preponderance of opinion in Europe appears to be that this is to be seen as a matter of partner nonparticipation or underparticipation. But other cases may be less clear-cut. Even if in-work poverty is construed as largely a problem of low-household work intensity, the question arises what can be deemed as sufficient level of work intensity. It is not self-evident that this is to equal all working age, work-capable adults in the household to be in full-time work the whole year round. Societal norms may differ across countries. In the Netherlands, for example, a four-fifths job per adult appears to be closer to the norm of full-work intensity. Also, household composition may be deemed to matter. It is not self-evident that a lone parent with young children is expected to work full-year, full-time before additional income support is to be considered legitimate if his or her earnings fall short of the poverty threshold.

Poverty is, to a large extent, far from exclusively associated with low-work intensity at the household level (see Corluy and Vandenbroucke, 2013; De Graaf-Zijl and Nolan, 2011). This brings into view a wide variety of potential policies that can help households to increase if not maximize their work intensity. These include policies aimed at boosting the demand for workers, and particularly the demand for people with low levels of education or weak work experience. Employer subsidies or reductions in employers' social security contributions are an example here. At the supply side, policy can stimulate (e.g., through fiscal reform) or support (e.g., through child care) people to take-up work or to increase working hours. What mix of policies will work best in a given context will depend on the composition of the low-work-intensity population and on the underlying causes of low-work intensity.

Yet, and this is crucial, it must be recognized that even if such policies succeeded in getting every single nonemployed person into work, or every household to a level of full-work intensity for that matter (and all empirical evidence to date suggests this to be highly unlikely), this would not guarantee the elimination of poverty. What policy can do to help households in these circumstances is again likely to depend on such factors as the institutional and policy context in place, labor market conditions, and the profile of the population in need of support.

In some EU countries, and certainly outside of the EU, minimum wages remain nonexistent or low relative to average wages, but in a range of others they do suffice to keep single persons reliant on them out of poverty. Thus, it would appear sensible for countries with nonexistent or very low minimum wages to contemplate introducing or increasing these. However, the route of introducing or boosting minimum wages to the upper ranges currently prevailing in Europe (relative to average earnings) would, even in the absence of negative employment effects, not be sufficient to eradicate in-work poverty. Even in countries where minimum wages are comparatively high they do not suffice to keep sole-breadwinner households out of poverty, especially when there are dependent others or children. Minimum wages have probably become inherently constrained in providing minimum income protection to sole-breadwinner households, especially in countries where relative poverty thresholds have become essentially determined by dual earner living standards.

For low-earnings households, only direct household income supplements may offer a reasonable prospect to a poverty-free existence, especially when there are dependent children. Such "in-work benefits" are now often associated with Anglo-Saxon-type "tax credits" such as the EITC in the United States and the WTC in the United Kingdom. It is increasingly argued that more effective redistribution will not come from augmenting/expanding the traditional channels of income support, for example, more generous social insurance or social assistance levels, or from higher minimum wages. These are seen not only as failing to address today's social risks and needs, but also as exacerbating underlying problems such as exclusion from the labor market and entrapment in passive benefit dependency. Worse, these are considered as standing in the way of innovative mechanisms of social protection that are proactive and self-sufficiency enhancing, such as active labor market policies and services such as child care and improved education and training.

The options to consider, then, are other forms of (targeted) income supplements for households that provide some level of income protection and that are also conducive to labor market participation. As Kenworthy (2011) puts it, "Given the importance of employment and working hours for the market incomes of low-end households, policy makers must guard against programs that provide attractive benefits without encouraging or requiring employment. An ideal transfer would be one that both boosts the incomes of low-earning households and promotes employment by able working-aged adults. As it happens such a program exists. Referred to variously as 'in-work benefit' or 'employment-conditional earnings subsidy', it is best exemplified by the Working Tax Credit (WTC) in the United Kingdom and the Earned Income Credit (EITC) in the United States" (p. 44).

Under these schemes households with low earnings do not pay taxes but instead they receive additional money through the tax system. In the United States, the 1993 expansion of the EITC created the country's preeminent antipoverty program for families of working age. The United Kingdom has also implemented and extended several schemes (and in fact did so earlier than the United States), culminating in the Universal Credit.

Clearly, Anglo-Saxon-style negative income taxes have been garnering increased interest of late. As Immervoll and Pearson (2009) note, "Even in the mid-1990s, twenty years after such schemes were first introduced in the United Kingdom and the United States, such schemes were seen as interesting but unusual [. . .] it seems reasonable to conclude that IWB schemes are now mainstream policies in many countries."

That is perhaps somewhat of an overstatement. Several European countries have contemplated introducing Anglo-Saxon-style tax credits, or have done so in some form. Examples here include the "Prime Pour l'Emploi" (PPE) and the Revenue de Solidarité Active (rSa) in France, the "Combination Credit" in the Netherlands, and a "Low Wage Tax Credit" in Belgium. Yet, the reality is that most of these schemes exhibit only a faint resemblance to the EITC or the WTC. Sweden has a scheme that goes by the same name in English as its U.S. counterpart, EITC. It was introduced in 2007, and was reinforced in 2008, 2009, and 2010. The stated motive for the reform was to boost employment; in particular, to provide incentives for individuals to go from unemployment to, at least, part-time work. The scheme is different from the U.S. scheme in that it is a nonrefundable tax credit. Also, because the tax unit in Sweden is the individual and not the household it works in effect as a tax relief on low individual earnings. In that respect it is similar to personal social security contributions relief measures elsewhere.

While tax-channeled in-work benefits targeted at households with low-earnings remain of limited significance in most European countries, it is of course the case that many countries have child benefit systems that provide an additional income to workers and their families (Van Mechelen and Bradshaw, 2013). Child benefits have generally lost ground. For a couple with two children, the size of the child benefits package, expressed as a percentage of the gross minimum wage, declined in the majority of countries awarding these benefits. For single parents with two children the trend was somewhat more favorable in a number of countries. The decline of child cash benefits, both in value as in their importance in net disposable income, is discussed more extensively in Van Mechelen and Bradshaw (2013). Interest in EITC type schemes remains strong, however, in the public debate and in the academic literature (Aaberge and Flood, 2013; Allègre and Jaehrling, 2011; Crettaz, 2011; Kenworthy, 2011; Marx et al., 2012a). This interest seems entirely legitimate. The empirical evidence shows the U.S. EITC, in combination with other policy reforms and several increases in the minimum wage, to have produced some significant results, including marked increases in labor market participation and declines in poverty among some segments of the population, especially single-parent households (Eissa and Hoynes, 2004; Hotz and Scholz, 2003). It needs to be noted, however, that these initial results occurred in favorable economic circumstances, including strong labor demand and low unemployment. The relatively strong increases in labor supply of single mothers in the U.S. setting also resulted from welfare reform—notably, the transformation of the social assistance scheme into a temporary support system with time limits on the duration of benefits. This clearly provided a strong push incentive, with the EITC acting as pull incentive. Not all who were forced out of passive dependence found their way to work (Grogger, 2003, 2004). In addition, as the survey by Holt (2011) reveals, there is considerable evidence of incomplete take-up (around 75% according to some estimates), although exact estimates are hampered by the fact that there is no systematic tracking.

There are potential downsides to subsidizing low-paid work. While EITC is intended to encourage work, EITC-induced increases in labor supply may drive wages down, shifting the intended transfer toward employers. Rothstein (2010) simulates the economic incidence of the EITC under a range of supply and demand elasticities and finds that in all scenarios a substantial portion of the intended transfer to low-income single mothers is captured by employers through reduced wages. The transfer to employers is borne in part by low-skill workers who are not themselves eligible for the EITC. There is some empirical evidence that corroborates the potential wage erosion effect of EITC (Chetty et al., 2013; Leigh, 2010).

Yet, whether EITC type schemes can work elsewhere, as Kenworthy (2011) and others suggest, is not self-evident. The sociodemographic make-up of the United States differs from that in most European countries; there are more single adult (and parent) households and also more multi-earner households. The dispersion in earnings is also much more compressed in most European countries, where, in addition, benefits are generally higher relative to wages (including minimum wages) and less subject to means-testing if they derive from social insurance. This also implies that benefit entitlements of household members are less interdependent, possibly weakening the potential impact on labor supply. Many countries have individual taxation, and the trend is moving away from joint taxation of couples.

In order to be effective as an antipoverty device and at the same time affordable within reasonable limits, such measures need to be strongly targeted. However, strong targeting at households with low earnings is bound to create mobility traps, which can only be avoided if taper-off rates are sufficiently flat. That comes at a very considerable cost if the lower end of the household earnings distribution is densely populated, as is the case in many European countries. This cost can only be avoided by making the amount of the tax credit itself smaller, but in that case the antipoverty effect is reduced. Simulations by Bargain and Orsini (2007) for Germany, France, and Finland, by Figari (2011) for four southern European countries (Italy, Spain, Portugal, and Greece) and by Marx et al. (2012a) for Belgium, shed doubt over the applicability of EITC type systems in other settings. In an earlier study, Bargain and Orsini (2007) investigated the effects on poverty of the hypothetical introduction of the British scheme (as it was in place in 1998) in Germany, France, and Finland, using EUROMOD for 2001. They found that the antipoverty effects of a U.K.-type tax credit (similar in design and relative overall spending) would be very small in these countries, especially relative to the budgetary cost. For Belgium, the hypothetical introduction of the United Kingdom's WTC is shown to yield a limited reduction in poverty at the cost of possible weakened work incentives for second earners (Marx et al., 2012a). Figari (2011) notes that the presence of extended families in southern Europe does not allow for such policies to be well targeted at the very poorest. Bargain and Orsini (2006) have concluded that "interest in such schemes is destined to fade away." Whether this is true remains uncertain and indeed doubtful, but EITC type negative tax credits are not obviously suitable for wholesale emulation throughout continental Europe. In Germany, for example, the labor market has undergone some profound changes over the past decade. Low-paid employment has become far more prevalent and in-work poverty seems to have increased. It is not unlikely that a simulation such as the one performed by Bargain and Orsini on 2001 data would yield different results today. A recent study by Giannelli et al. (2013) analyzes the quality of new jobs created in Germany between 1998 and 2010 and find that the reforms of the 2000s (Hartz reforms) reinforced an existing trend of increasing wage inequality and lower wages among the least advantaged individuals. Although, as found by Card et al. (2013), a great deal of the increase of wage inequality in Germany for the period 1985–2009 is due to the increasing heterogeneity in job premiums and the raise of assortativeness in the matching between workers and establishments.

Clearly, simulations demonstrate that in-work benefit schemes that work well in certain settings do not necessarily perform equally well in a different context. Family composition, individual earnings distributions, and family income structures drive outcomes in a very substantial way. It remains to be explored whether alternative designs are conceivable that have better outcomes in continental European settings and that are realistically affordable.

23.2.5 Pensions

The terminology "pillars" is widely employed (Holzmann and Hinz, 2005) to capture the different elements of pension systems, as they operate within, for example, Bismarckian or Beveridgean welfare states. Bovenberg and Van Ewijk (2011) offer a typology of four models of pension systems based on the dimensions of governance (private vs. public) and individual choice (mandatory vs. voluntary), which are related to the classification of welfare states by Esping-Andersen (1990). As pension systems in rich economies have, simultaneously or not, characteristics of social insurance and poverty prevention, and different forms to finance benefits, a more flexible taxonomy of pension systems is used by the OECD (see Figure 23.6).

There are three main visible tiers forming the retirement-income system. The first one is intended to prevent old-age poverty and is publicly financed. Within this tier there are basic benefits paid at a flat rate, resources-tested (means and assets) benefits, and minimum pensions. The second tier is composed by mandatory schemes that can be public or

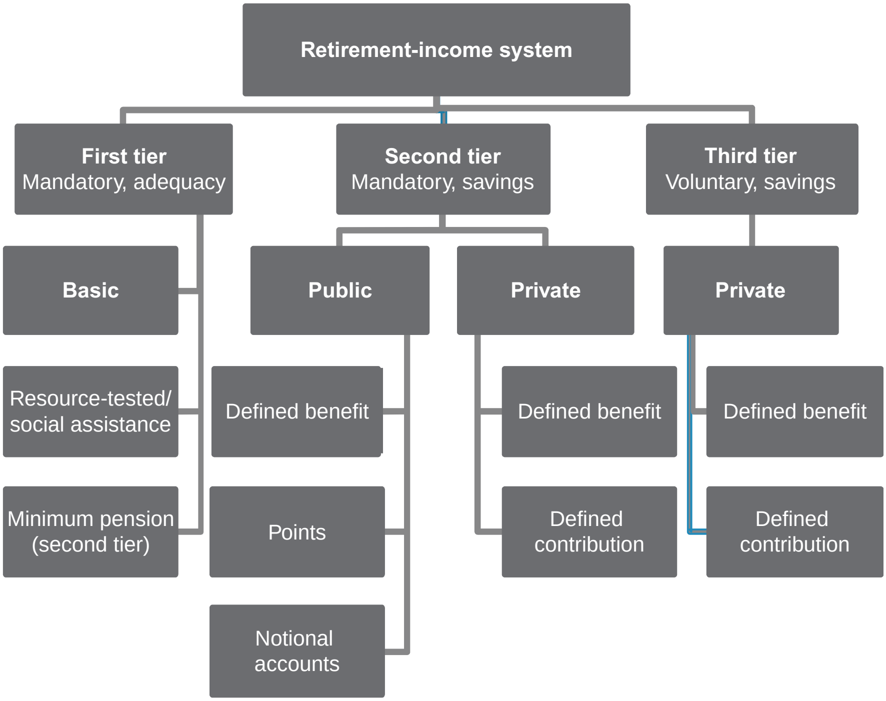


Figure 23.6 Taxonomy of different types of retirement-income systems. *Source:* OECD (2011b).

private. The public schemes offer defined benefits (DB) where the pension entitled is a function of individual contributed years and income. A system of points earned with each year income and accrued up to retirement age is also possible (e.g., occupational plans in France). A third plan under the public provision of the second tier is the Notional Defined Contributions, which is used in Italy, Norway, Poland, and Sweden. Under this plan, the individual contributions are recorded by the pension institution and offered a return rate. Once the retirement age is reached, such contributions are converted into pensions through an actuarial formula. The second tier also includes compulsory private (occupational) managed pensions, which can be DB or defined contribution (DC) types. Finally, the third tier is composed by voluntary private plans.

The composition of each plan within and between countries varies to a great extent. From 34 OECD countries, 14 have mandatory private schemes, 12 have public resources-tested benefits, 13 have basic flat rate benefits, and 18 have minimum pensions. Furthermore, DB pensions are present in 20 countries while DC pensions exist in 11 economies. For more details about the composition of pension plans by county, see section II.1 of OECD (2011a).

The adequacy of pension benefits is broadly measured by the replacement rate—namely, the ratio between pensions and average wages. Figure 23.7 reports the net

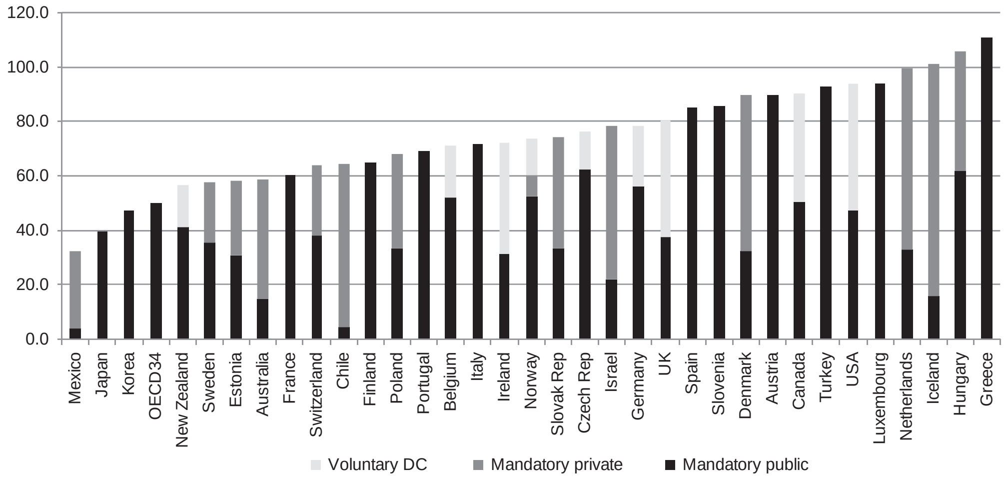


Figure 23.7 Net pension replacement rates by pension schemes in OECD countries. *Source: OECD (2011b). Authors' elaboration.*

pension replacement ratios in OECD countries with data from late 2000s. On average, the replacement rate is 50% in mandatory public plans while it is 43% in mandatory private plans, and 28% in voluntary plans. Overall, the mandatory systems show a replacement rate of 68%, which rises to 77% when voluntary plans are added. Furthermore, one can observe that adequacy differs significantly among countries and pension schemes. For example, in Japan, Korea, and Mexico the overall net replacement rate is lower than 50% while in 13 of 34 countries this figure is above 80%. All of the replacement ratio figures are lower when gross income and pensions are considered because income taxation burden for retirees is milder than for the working population. The mandatory systems have a gross replacement ratio of 57% and this reaches 64% when voluntary plans are included.

Typically, individuals at the beginning and the end of the life cycle face higher poverty rates. This U-shaped relationship by age groups has been maintained during the last decades, but the poverty rates have shifted impressively in favor of the elderly and in detriment of children and the young. Figure 23.8 from the OECD's *Unequal Growing?* shows clearly the sharp reduction of poverty risk for old-age individuals between the 1970s and the 2000s in OECD countries. Moreover, women report more poverty rates than men. The poverty gap by gender significantly increases for older ages. As explained in OECD (2008), Smeeding and Sandstrom (2005), and Vignoli and De Santis (2010), the risk of living in poverty is higher for elderly women because they have gained less pension rights during their working life, and they are more likely to live alone after the death of their spouses. In this regard, studies from Burtless (2009) and Vignoli and De Santis (2010) alert on the trends of new living arrangements (shrinking of the household size of the elderly) that jeopardize the living conditions of the elderly and increase the risk of falling into poverty. As a feedback mechanism, the larger participation of the elderly in pensions

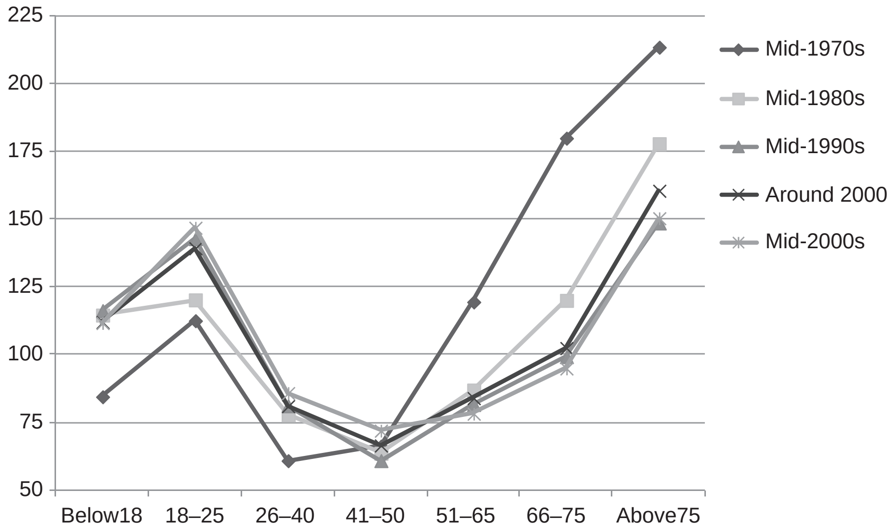


Figure 23.8 Risk of relative poverty by age of individuals in OECD-7 countries (poverty rate of the entire population in each year = 100). *Notes: Relative poverty risk is the age-specific poverty rate divided by the poverty rate for the entire population times 100. The poverty threshold is set at 50% of median income of the entire population. OECD-7 is the average for Canada, Finland, Greece, the Netherlands, Sweden, the United Kingdom, and the United States. Source: OECD (2008).*

and transfers will allow them to live alone without the need of other relatives, increasing in this way the risk of poverty. McGarry and Davenport (1998) are also aware of the effects of survivorship benefits for pensions on the poverty rates of U.S. widows given the scarcity of pension wealth of women.

The role of pensions in reducing poverty is particularly important due to the large share of old-age income coming from social security. On average, public transfers (earnings-related pensions, resource-tested benefits, etc.) to people over 65 during the mid-2000s represent 60% of their incomes. In some countries this figure reaches 80% or more (France, Hungary, Slovak Republic, and Belgium). The rest of the sources are divided in-work income (21%) and capital income (19%). The shares of incomes from work are large in Japan, Turkey, Mexico, and Korea where those represent about 50%. The average share of capital income for the elderly in Netherlands, United Kingdom, Switzerland, Canada, Australia, Denmark, and the United States is about 41%. Note that occupational plans are included in capital income sources, so that this component includes pension incomes. All these figures by country can be consulted in OECD (2011a). It is also observed that the reduction of market income poverty attained with transfers and taxes is greater for old-age people than for working-age people (OECD, 2008).

A number of recent studies have sought to measure the contribution of pensions in reducing old-age poverty across countries. Smeeding and Williamson (2001) use LIS data to estimate the effect of public pensions on poverty rates of the elderly in eight developed economies for mid-1990s: Australia, Canada, France, Germany, the Netherlands, Sweden, the United Kingdom, and the United States. The poverty rate for old-age people would be 84% in average if only market income is considered. This is reduced to 71.8% when occupational pensions are added, and then this falls to 21.2% when universal and social incomes are included. With social safety net transfers, the average poverty rate drops up to 13.2%. Different from the English-speaking countries, the greatest redistributive effects are found in France, Germany, the Netherlands, and Sweden. Similar trends are found when Smeeding and Sandstrom (2005) analyze data for the early 2000s. In both works, it is found that pensions are more effective to reduce poverty of old-age males than in old-age females. With early 2000s data, pensions, income social transfers, and safety net transfers reduce poverty of elderly women up to 24.3% while that figure is reduced up to 13.3% for both sexes. As women participate to less extent in the pension system, the safety net transfers are more important for them to reduce their risk of poverty, and the contrary holds for occupational pensions in the case of men. In a similar exercise by Lefebvre (2007), it is found that poverty alleviation due to pensions is less effective for the very old (75+) than for the old (65–74). This feature combined with gender depicts a very negative picture for very old women, who in turn, represent the majority of members in the oldest old cohort. Moreover, micro-simulation models like the one implemented by Dang et al. (2006) arrive at similar conclusions. Chapter 24 of this book shows other relevant micro-simulation models.

There is concern about the sustainability of public pension expenditures due to the accelerated aging process in developed economies; and, in particular, there is a legitimate worry about the effects of the reforms aiming to attenuate it on old-age poverty and inequality (Arza and Kohli, 2008; Börsch-Supan, 2012; Burtless, 2006). Although, as indicated in the reports by Zaidi et al. (2006a,b) the pension reforms promoted by the World Bank were mainly driven by financial sustainability issues, and little concern was put on the effects on the living standards of the retirees. These reports offer an important effort to estimate the long-term effects of a variety of pension reforms in EU countries—undertaken between the 1990s and the 2000s—on the poverty and living standards of the elderly. In Zaidi et al. (2006b), the authors find a strong negative relationship between the generosity of public pensions and the at-risk-of-poverty rates among the 65 and older, and they foresee a decline of the pension generosity (for years 2025 and 2050) on the basis of the analysis of each pension reform. These two combined findings will result in an increase of the poverty rate for the vast majority of countries analyzed (Estonia, Malta, Austria, Italy, Belgium, Denmark, Spain, France, Latvia, Lithuania, Portugal, Slovenia, Finland, and Sweden). Only Ireland and Cyprus appear to have a reduction of the 65 and older poverty rates. However, as warned by the authors, these results have to be taken with caution as no behavioral responses are considered.

In a more static framework, Van Vliet et al. (2012) estimate the effects of pension reform on poverty and inequality in European countries. They acknowledge that recent shifts from public to private provision in pensions are still limited in Europe but that this is important for some countries. They estimate the effects of those changes on old-age inequality and poverty with OLS panel data regressions, but they do not find substantial effects on those variables. Nonetheless, they cast the limitation of their analysis by indicating that the reforms may be affecting only to new and future retirees.

Looking at the effects of public transfers and taxes in a more general way, some authors appeal for a rebalance of the spending from pension programs toward programs aimed to prime-age people and their children at the bottom of the income scale, which could reduce poverty rates to a greater extent (OECD, 2008). As pointed out by Dang et al. (2006), social protection systems are very old-age oriented in the EU with the elderly receiving much more cash transfers than the working population. They show that even high old-age spending countries can leave significant pockets of elderly in poverty while others with lower expenditures in old-age can be more successful at limiting the risk of poverty. Furthermore, their simulations indicate that there is scope to reorient the expenditures from old-age to working population and rebalance the tax liabilities in favor of the working population. These changes will not jeopardize the living standards of the elderly if the reforms include proper safety net measures.

The role of public pensions in reducing inequality can be very large because these pensions represent a large fraction on income in old-age. During the mid-2000s, public cash benefits accounted for 70% of income of retirement-age individuals in 24 OECD countries, and in many of them the figure was above 80% (OECD, 2008). In countries where public pensions are important, the effect of re-ranking when one uses the distribution of market or disposable income as the counterfactual can be large. Mahler and Jesuit (2006) find a sizeable effect of pensions (public and private) in reducing the Gini coefficient on 13 rich countries during the period 1980–2000. On average, the Gini is reduced from 0.43 to 0.27 when all taxes and transfers are considered, with a reduction of 0.039 points from taxes and 0.121 from transfers of which 0.068 comes from pensions, 0.013 from unemployment transfers, and 0.040 from other transfers. In Belgium, Sweden, and France, the reduction in the Gini is about 0.10 points due to pensions, while in the United States, Canada, and Australia it is only about 0.04 points. Lefebvre (2007) computes the marginal contribution of earnings, property income, private pensions, and public transfers on total inequality in 19 EU countries. It is found that public pensions decrease inequality in all countries and that private pensions increase inequality in all countries except in Ireland and France. Similarly, Caminada et al. (2012) disentangle the changes of contributions of different income components in reducing inequality between the mid-1980s and mid-2000s in 12 LIS countries. Around 1985, the primary income Gini falls 0.139 points after transfers and taxes, while around 2005 this drop is about 0.163 points. The authors estimate that this increase in redistribution is mainly due to the state old-age and survivor benefits, which account for 60% of the total change. Different designs of pension systems have diverse effects on inequality. For example, Benedict and Shaw (1995) with data from the early 1980s, find that private pensions in the United States increase inequality among unionized workers by 21% with respect to observed wage inequality. On the reforms undertaken in Europe since the mid-1990s, Van Vliet et al. (2012) do not find evidence of important effects of those reforms on income inequality.

In general, the assessment of inequality is made in one single year, but studies such as Burtless (2006) emphasize that this approach can overestimate the redistributive impact of pensions. This is related to the question of what is the proper counterfactual distribution to use when one analyzes the impact of pensions. If pensions are simply absent, it is expected that individuals will look for other forms of savings to afford their old-age. Different living arrangements can also be different if pensions would be nonexistent or less generous, which will cause other redistributive effects (Burtless, 2006, 2009). In this regard, some authors favor the estimation of the distribution of lifetime income (e.g., Deaton et al., 2002; Liebman, 2002) although the data requirements are more demanding. This approach shares features with a growing literature studying lifetime income inequality (see for instance Aaberge and Mogstad, 2012) which highlights a life-cycle bias that overestimates income inequality when only one or a few years are analyzed.

23.3. BEYOND SOCIAL PROTECTION

23.3.1 Noncash Social Spending and Poverty

While cash transfers form a substantial proportion of overall social expenditure and have a pronounced impact on household incomes and poverty, other forms of social expenditure—such as health, housing, and perhaps education (which is sometimes included as "social" spending and sometimes not)—may also have substantial direct and indirect effects. Table 23.4 shows spending on cash transfers and on other forms of social expenditure—which one can think of as benefits in kind from a household perspective—based on the OECD's social expenditure database before the onset of the economic crisis in 2007–2008, which has boosted expenditure on cash transfers in many countries. This shows that in about half the countries shown, cash transfers significantly outweighed such benefits in kind—notably, in the "continental/corporatist" countries like France, Germany, Belgium, and Luxembourg; in the southern countries of Italy, Spain, and Greece; and in Poland. However, in many of the other OECD countries, overall social spending was fairly evenly balanced between cash transfers and other spending. This is using a definition of social expenditure that does not include education, so if one adds education spending, the relative importance of noncash spending is even more obvious, as brought out in Marical et al. (2008). They conclude that public spending on health, education, and "other services" in the OECD social expenditure database represents an amount comparable to public cash transfers, exceeding those transfers in 11 OECD countries.

The impact of such noncash spending on poverty is difficult to assess for various reasons (see for example Currie and Gahvari, 2008; Garfinkel et al., 2006). One approach

Table 23.4 Social expenditure distinguishing cash and noncash benefits as percentage of GDP in OECD countries, mid-2000s

Country	Cash transfers % of GDP	Noncash social benefits % of GDP
Australia	8.1	6.7
Austria	18.4	8.2
Belgium	16.2	9.1
Canada	8.8	9.4
Czech Republic	11.4	7.8
Denmark	13.8	11.8
Finland	15.3	9.9
France	17.5	10.8
Germany	15.9	9.9
Greece	13.4	7.1
Hungary	13.8	8.7
Ireland	8.4	7.7
Italy	16.7	7.7
Japan	10.2	8.1
Luxembourg	13.9	8.8
Netherlands	11.1	8.5
New Zealand	9.7	8.4
Norway	10.9	10.1
Poland	15.7	4.9
Slovak Republic	10.2	6.1
Spain	13.1	7.4
Sweden	14.5	13.6
Switzerland	11.8	7.8
UK	10.3	10.5
USA	8.0	7.0

Source: OECD social expenditure database.

employed in comparative studies (Callan et al., 2008; Marical et al., 2008; Paulus et al., 2010; Smeeding et al., 1993) and in national studies (Aaberge and Langørgen, 2006; Callan and Keane, 2009; Harding et al., 2006; Nolan and Russell, 2001; Wolff and Zacharias, 2007) is to use micro data to assess who is benefiting from such expenditure and to what extent, and to compare overall inequality and (sometimes) poverty levels when this noncash income is included. With some studies this also means allocating indirect taxes to households and deducting them to arrive at a "final" income concept. Major decisions have to be made about how to value the benefits to users of services as has been debated in the literature for three decades. The empirical studies have shown that these can have a marked impact on the measured outcomes—notably, in the case of health spending where particularly challenging conceptual issues have to be addressed.

One complication is that services, which in principle are provided free or in subsidized fashion to everyone, may actually be readily available only in certain areas or to certain groups, or even if available may be taken up to a varying degree by those with higher versus lower levels of income or education. Information on actual use patterns may not always be available, and attributing a common value across a particular age group, for example, may be misleading. Empirical studies thus make use, where possible, of information—generally from household surveys—of actual usage patterns for the range of services involved, but this may not cover all the areas of expenditure one wants to include.

Difficulties then arise, though, first of all because one does not know whether households would have bought the same amount of the goods or services in question if those were not provided free or at a subsidized rate. Recipients may place a value on noncash benefits that is less than what they would have to pay for the good or service in the market, because the recipient has no choice in its allocation. However, a U.S. study of food stamps suggested that where the item is a basic necessity and the in-kind transfer is smaller than the amount the household would normally spend on that good, the value to the recipient may be very close to the market price (Moffit, 1989). Unlike food, what is meant by market price for many of the services provided by the state may itself be unclear since they are not available in the market—the most obvious examples being defense or law enforcement. If one takes the supply price (i.e., the cost to government) as the point of reference, the optimal level of provision will equate the marginal benefit with this price times the marginal cost of public funds. In any case, the widely used approach in empirical studies is simply to assume that the value of a particular (unit of a) service is equal to the average cost of producing it. Use of such an average may mask variations in quality of the service provided to different socioeconomic groups—for example, in the quality of the health care provided to the rich versus the poor—and that is another important aspect that is very difficult to capture empirically.

The second general issue arises where the noncash benefit covers something like health care, which is required to meet a specific contingency affecting only some households in a given year. In those circumstances, if we simply add the cost of the free or subsidized services supplied to the households consuming them, sick people will be richer than the healthy at any cash income level. One can in those circumstances attempt to also take the additional "needs" of such persons into account by elaboration of the equivalence scales employed—drawing on, for example, recent studies focused on the costs associated with disability such as in Jones and O'Donnell (1995) and Zaidi and Burchardt (2005)—but this remains underdeveloped. A more widely employed approach is that instead of basing values on the household's own consumption, one attributes to all those eligible for state provision an extra income equal to the insurance premium they would have to pay to obtain the same level of cover in the market. Even assuming the cost of this cover can be established satisfactorily, a serious problem remains. Even the insurance value could be worth enough by itself to bring a household above the poverty threshold when it might still have insufficient cash income to buy enough food, clothing, or shelter, reinforcing the point that the in-kind transfer does not represent command over resources in the same way that cash income does. Furthermore, even with the insurance approach, the fact that different households have different underlying needs should be taken into account in arriving at conclusions about the welfare implications of in-kind benefits (see Aaberge et al., 2010).

The final, and fundamental, issue to be noted relates to the time-period employed. In measuring poverty and income inequality annual income is most often the focus, but in thinking about the consumption of education or health care and the value of the in-kind benefit they represent it would be natural to take a life-cycle approach, since the benefits are often long-term rather than confined to the point of use. Such an approach is very demanding in data terms and involves a wide range of assumptions for which it is difficult to find a robust empirical basis.

The results of recent empirical studies on this topic are of significant interest in the broader context of welfare state institutions and policies and their impact on poverty. Marical et al. (2006—and also chapter 9 in OECD, 2008) look at the impact of public spending on health, education, and social housing on income inequality in OECD countries, concluding that they generally contribute to narrowing inequality, though not usually by as much as cash transfers and direct taxes combined; they do not look at corresponding results for poverty. Paulus et al. (2010) on the other hand assess the impact of valuing noncash or in-kind benefits from public housing subsidies, education, and health care in five European countries, recalculating both inequality and relative poverty measures when this value is added to cash income. In such an exercise, the relative income poverty threshold—in this case 60% of median equivalized income—is itself recalculated, rising by between about one-fifth and one-third in value when in-kind benefits are included. The proportion of persons falling below that threshold is found to be much lower than the corresponding figure based on cash income in all five countries, with reduction being greatest in the United Kingdom where the poverty rate falls by half and least in Greece where it still falls by one-third. There are also major effects on the composition of those falling below the threshold, with the reduction in poverty rate greatest for children and older people (since the incidence of spending on education and health care is particularly concentrated on them). This pattern is familiar, having featured strongly in Smeeding et al.'s (1993) early comparative study covering seven countries based on data in LIS.

Sutherland and co-authors caution that "it is doubtful whether these results should be interpreted as having any bearing on the assessment of poverty or inequality from a welfare perspective" (p. 259), being mainly of interest in showing the scale of noncash incomes relative to cash incomes, but without taking into account the needs of individuals for health care or education. The study goes on to attempt to take the variation in those needs into account via modifying the equivalence scales employed. It finds that the distributional effects of noncash transfers on several summary income inequality measures are then far more modest; corresponding results for poverty rates are not reported, but it seems likely that the same would be true in that case. It is also worth highlighting the argument by Bourguignon and Rogers (2007) that once the intertemporal or intergenerational nature of the effects of many social expenditures are recognized, it is no longer possible to assume that they are equivalent to cash transfers, food subsidies, and other programs of direct redistribution. Education spending is an investment in future generations and may have redistributive effects for these generations, but may worsen distribution initially. Moral hazard makes it infeasible to borrow against the human capital of one's descendants, so an increase in public education expenditures financed by an increase in a neutral tax may actually be regressive for the generations with school-age children. Poor households in this generation pay the tax and receive no benefit, whereas rich households pay the tax but may recover it through intergenerational reallocation of consumption (that is, smaller bequests to their children). Intergenerational accounting may then be necessary to more fully capture the redistributive and poverty-related effects.

23.3.2 The Labor Market, Education, and Active Labor Market Policy

Income derived from the labor market is central to the overall distribution of income and to poverty and disadvantage at household level (see for example OECD, 2008). Even for those not currently earning (via employment or self-employment), previous labor market experience may determine current entitlement to social protection or to occupational pensions. A wide variety of studies on poverty in individual countries, both descriptive and econometric, find that those in work are much less likely to be poor than the unemployed or working-age inactive. Cross-country differences in labor market performance and structure then seem a natural starting point in seeking to understand cross-country variation in poverty rates (Burniaux et al., 1998; Förster and Mira d'Ercole, 2005). The poverty rate among the working-age population varies greatly across OECD countries and is indeed the main contributor to overall poverty headcounts (see for example OECD, 2009). However, at the country level working-age poverty—overall or for specific groups—is not in fact strongly linked to employment rates. Burniaux et al. (2006) report some relationship between female participation rates and poverty rates across OECD countries, but it is not particularly strong. Poverty rates are generally lower in low unemployment countries and vice versa, but there are notable exceptions. High employment rate is not a sufficient condition for low poverty among the working-aged population (see also the simulations in Marx et al., 2012b). At the aggregate level, then, employment performances are not the main driver of cross-country differences in the overall poverty risk among the working-age population (OECD, 2009).

There is thus a contrast between micro studies on poverty and the labor market in individual countries, which tends to focus on the labor market situation and experience of individuals and their households and the characteristics associated with good rather than bad labor market outcomes for them, and comparative studies at the aggregate level, which focus on labor market institutions and performance. The relationship between individual characteristics and labor market outcomes is of course a core concern of labor market research, as is the structure of earnings in terms of overall dispersion and differentials. (For reviews see the Handbook of Labour Economics; Ashenfelter and Card, 1999, 2011; Ashenfelter and Layard, 1987; Blau and Kahn, 2008; and Chapter 18 of this volume by Checchi and Salverda.) The extent to which individual disadvantage and relatively bad labor market outcomes manifest themselves in high-poverty rates depends on the household, labor market, and institutional settings in which those disadvantages are experienced. Comparative studies of the relationship between poverty and the labor market at the aggregate level include collective bargaining structures, the role of unions, minimum wages, and so forth, in the explanatory variables employed as key aspects of labor market institutions (see, for example, Burniaux and Mira d'Ercole, 2006). These may often be embedded in wider sets of variables covering, for example, welfare spending and structures, intended not only to serve as controls but also to capture broader concepts of the welfare state "regime," as discussed in Section 23.1.4. This reflects a recognition that labor market institutions, while central, are inextricably bound up with the broader welfare state, and that the impact on poverty of, for example, a minimum wage will vary depending on that broader context—as brought out in our discussion of in-work poverty and social protection transfers in Section 23.2.4.

A core element of that broader welfare state, strongly linked to the labor market, is the education system and educational spending. Once again a contrast may be drawn between micro studies on the relationship between educational attainment, earnings, and poverty at individual or household level, and studies at the aggregate level that focus on the education system and spending and their impact on economic performance and poverty. The relationship between educational attainment and earnings/labor market outcomes for individuals has been a major preoccupation of labor market research since the earnings equation first derived by Mincer (1958) became a basic tool of analysis, but the broader role of education as a facilitator or engine of economic growth is also a major focus of research. The concept of "human capital" has become embedded since the "Chicago School" of economics (see especially Becker, 1964; Mincer, 1958), with human capital seen as similar to physical means of production in that investment in enhancing capacities and skills, notably through education and training, also increases future productive capacity. Microeconomic investigation of this relationship via estimation of the returns accruing to the individual in terms of earnings is the topic of a vast array of empirical economic research, including investigation of the extent to which the positive earnings differentials for the more educated may be interpreted as a causal impact of education itself rather than selection (on which see for example Card, 1999; Machin, 2008). The impact of educational attainment on the likelihood of being in poverty is also a consistent finding from microeconometric analysis of individual OECD countries and holds whether poverty is measured in terms of low annual income, persistent low income, or levels of deprivation (see for example Fouarge and Layte, 2005; Layte and Whelan, 2002), though the relative and absolute "penalty" paid for low educational attainment in terms of enhanced poverty risk varies substantially across countries.

The implications of this individual-level link between educational attainment and poverty risk for aggregate performance and for policy is not as straightforward as it is often taken to be, and requires further research. Improving the education and skills of the workforce has assumed a central role in strategies to promote economic growth and tackle poverty and exclusion. This is illustrated by the European Union's, 2013 *Social Investment Package*, which focuses on policies designed to strengthen people's skills and capacities, including education and child care as well as active labor market policies (see European Commission, 2013), or in a U.S. context by the Obama administration highlighting that "To prepare Americans for the jobs of the future and help restore middle-class security, we have to out-educate the world and that starts with a strong school system."[5] This reflects, in particular, the concern that the low-skilled in advanced economies are being left behind by rapid technological change in a globalized world economy, as discussed in depth in Freeman (2008) and Chapter 20 by Kanbur in this volume. On the role of education in this context, OECD (2011a) concludes that between the mid-1980s and the mid-2000s the sizeable disequalizing effect on earnings of factors such as technological change, more flexible labor market regulation, and less generous unemployment insurance was largely offset by growth in average educational attainment, up-skilling serving to reduce wage dispersion among workers and increase employment rates.

However, the corollary is not that continued expansion in education *per se* will be effective as an equalizing or antipoverty policy. As Checchi et al. (2014) emphasize, increasing average levels of educational attainment was associated with reducing dispersion in attainment in many OECD countries over the twentieth century, but with completion rates at second level approaching saturation in many rich countries, the main issue facing educational policies in most OECD countries now is whether they should pursue further expansion at tertiary level. Such expansion, depending on how it is brought about and underpinned, may not benefit those from poorer backgrounds, as we discuss in the context of intergenerational transmission of disadvantage in the next section. Research on how best to enhance skills in the middle and bottom parts of the distribution in secondary school, including performance in mathematics and languages, as well as issues of

[5] http://www.whitehouse.gov/issues/education, downloaded July 25, 2013.

school system structures, tracking, and early childhood education, discussed in the next section are thus also central to the research agenda from a poverty perspective.

Training and skill enhancement, as well as matching, are important components of the active labor market programs and activation strategies that are now widely seen as at the core of antipoverty policies (see European Commission, 2013; OECD, 2009). These have been the subject of a very substantial research literature, covering the evaluation of the impact of specific interventions and of active labor market policies more broadly. For reviews see Heckman et al. (1999), OECD (2005, 2007b), Card et al. (2010), and Kluve (2010). The general thrust of these evaluations, when carried out rigorously, was not particularly positive for a time, as reflected in Richard Freeman's summary that "Random assignment social experiments analysed with care .... have shown us that one favourite solution to labour market problems—training and other active labour market measures—have at best only modest effects on outcomes" (Freeman, 1998, p. 16). More recent evaluations have been more positive in tone, with OECD (2009), for example, concluding that activation programs can have a significant impact on unemployment. Card et al.'s (2010) meta-analysis of microeconometric evaluations yields particularly interesting findings from both a substantive and methodological point of view. They find subsidized public sector employment programs to have the least favorable impact estimates, whereas job-search assistance programs have relatively favorable short-run impacts, classroom and on-the-job training programs tend to show better outcomes in the medium-run than in the short-run, and programs for youths are less likely to yield positive impacts than untargeted programs. Methodologically, they find that—controlling for the outcome measure and the type of program and participants—experimental and nonexperimental studies have similar impact estimates, suggesting that the research designs used in recent nonexperimental evaluations are unbiased. They also note that the outcome variable used to measure program effectiveness matters, with evaluations based on registered unemployment durations being more likely to show favorable short-term impacts. The outcome variable is also clearly very important from a poverty perspective: It cannot be taken for granted that success in terms of a transition from unemployment into employment, even if sustained, leads to an escape from poverty because not all those benefiting may have been in poverty when unemployed. For those who were in poverty, the increase in income involved after taxes and withdrawal of benefits may not suffice to lift the household above a poverty threshold, as discussed at some length in Section 23.3. The rigorous evaluation of active labor market programs in terms of their impact on poverty remains a major gap to be filled.

As is noted in Card et al.'s (2010) work, active labor market programs are widely diverse. An effort to categorize these policies in relation to their political determinants is made by Bonoli (2010) on the basis of national variation across OECD economies. However, Bonoli found little regularity over time in these determinants, with a mix of leftist and centrist political parties in each period advocating active labor market policies. Moreover, Bruno and Rovelli (2010) compare and document differences in labor market policies in EU countries in 2000s and find that, in general, higher rates of employment are associated with more expenditure on active labor market programs for countries with a larger share of the population embracing pro-work attitudes. Recently, an OECD (2013) study analyzing activation programs in OECD countries and with more detail in Ireland, the United Kingdom, Japan, Norway, Finland, Switzerland, and Australia brings out the different responses of expenditures on activation programs after the economic crisis, finding it difficult to establish a common pattern.

23.3.3 Intergenerational Transmission, Childhood, and Neighborhoods

The intergenerational transmission of poverty and disadvantage continues to be a core concern for research and policy. Research on income mobility across the distribution is the topic of Chapter 10 by Jäntti and Jenkins, but here it is important to reiterate that there is substantial evidence from country-specific studies that mobility is particularly limited toward the bottom of the socioeconomic hierarchy, so that poverty is to a significant degree inherited across generations. Examples from research in the United States include Wilson (1987), Gottschalk et al. (1994), Duncan et al. (1994, 1998), Duncan and Brooks-Gunn (1997), and Corcoran (2001); for Canada, see Corak (2001); for recent U.K. studies include Sigle-Rushton (2004) and Blanden and Gibbons (2006), and similar studies that trace current poverty or disadvantage to conditions in childhood that exist for other rich countries. The likelihood of being a welfare recipient is also seen to be associated across generations—see, for example, Corak (2004) for Sweden and Canada and Page (2004) for the United States.

OECD (2009) concludes that variation in the strength of transmission of poverty across countries cannot reliably be assessed with the available evidence. However, the findings of Jäntti et al. (2006) showing considerably greater upward mobility in individual earnings from the bottom quintile in the Scandinavian countries than in the United Kingdom and especially the United States, and those of Raaum et al. (2007) that the intergenerational transmission of family earnings is also significantly stronger in the United Kingdom and even more so the United States than in the Scandinavian countries, are suggestive (see also Aaberge et al., 2002). Furthermore, recent studies by Esping-Andersen and Wagner (2010) and Whelan et al. (2013) have been able to exploit the availability of harmonized data from a special module on intergenerational transmission attached to EU-SILC in 2005. Esping-Andersen and Wagner estimate the impact of economic hardship during childhood on both educational attainment and adult income (controlling *inter alia* for lone motherhood and parents' education) in Denmark, Norway, France, Italy, Spain, and the United Kingdom. They conclude that economic hardship in childhood has no direct effects on adult income in any of the countries, but it does have powerful indirect effects via children's educational attainment; this effect disappears among the youngest cohorts in both Denmark and Norway but not in the other countries, leading the authors to conclude that the Scandinavian countries are more recently succeeding in minimizing the adverse consequences of economic want in childhood. This is consistent with Whelan, Nolan, and Maitre's study, which included a broader range of EU countries and found that factors such as parental class, parental education, and childhood economic circumstances/hardship had less influence on income poverty and a broader, multidimensional measure of vulnerability in social democratic countries than in countries in the liberal and southern European welfare regimes.

Understanding the mechanisms at work is clearly vital in designing strategies aimed at reducing the extent to which poverty is handed down from one generation to the next, and both causal channels and policy responses have been the subject of substantial bodies of literature (for reviews see D'Addio, 2007; Esping-Andersen, 2004a,b; Nolan et al., 2011). Studies focused on the United States show that the inheritance of poverty is connected with substantially less schooling (on average, poor children will have 2 years fewer schooling than nonpoor children), poor health, and crime (Duncan and Brooks-Gunn, 1997; Mayer, 1997), and similar if less dramatic effects have been documented for the United Kingdom (Gregg et al., 1999) and France (CERC, 2004; Maurin, 2002). Gregg et al.'s (1999) study controls for the child's abilities (via cognitive test scores at age seven), and still finds strong poverty effects. U.S. and British studies demonstrate strong negative effects of lone motherhood on child outcomes, but also suggest that the main reason has to do with poor economic conditions (Biblarz and Raftery, 1999; Gregg et al., 1999; McLanahan and Sandefur, 1994), while selection into lone parenthood may also be a factor (Piketty, 2003). Interestingly, Esping-Andersen and Wagner's (2010) multicountry study found no significant effects of lone motherhood on educational attainment or adult income having controlled for mother's education and childhood financial hardship.

The impact of genes/nature versus nurture and the interactions between them have been the topic of much debate in the broader intergenerational mobility literature, as discussed in Jäntti and Jenkins' Chapter 10. (See also Chapter 18 by O'Donnell, Van Doorslaer and Van Ourti for a detailed discussion on health and inequality.) From the point of view of transmission of poverty and disadvantage, the key thrust of recent findings is that cognitive skills and family finances matter, but so do noncognitive abilities, social skills, cultural resources, motivation and, more generally, the familial "learning milieu." Cognitive and noncognitive skills are influenced by family endowments that are neither strictly financial or genetic. Heckman and Lochner (2000) and Carneiro and Heckman (2003) have been influential studies, with their "learning-begets-learning" model stressing the fundamental causal importance of conditions in the preschool years, especially those related to behavioral and cognitive development. There is growing consensus in the literature that conditions when children are under age 6, or even 3, are decisive for their cognitive skills, sense of security, and ability and motivation to learn (Danziger and Waldfogel, 2000; Duncan and Brooks-Gunn, 1997). Substantial differences in children's cognitive abilities by parents' socioeconomic status emerge at early ages and carry through to subsequent achievements in education and earnings (e.g., Cunha and Heckman, 2007); poverty in early childhood has strong adverse effects on these later outcomes, partly because of parental traits such as poor cognitive and noncognitive skills and the effects of family "culture," in particular in terms of how it influences parenting behavior and child stimulation (de Graaf et al., 2000; Esping-Andersen, 2007).

This has significantly influenced thinking about the role of education in seeking to reduce intergenerational transmission of poverty. Mounting evidence suggests that differences in the design and financing of education systems *per se* seem to matter rather less than had been thought. There appears to be a broad consensus that early tracking according to ability reduces educational mobility across generations (see Hanushek and Woessmann, 2006), with the abolition of early tracking and the introduction of comprehensive school systems seen to have boosted educational attainment among the least privileged social strata in Sweden, Finland, and Norway. Since these are also countries in which welfare state redistribution increased substantially over the same period, it is difficult to identify how much of it was education reform or income equalization that produced higher mobility. However, Blanden et al.'s (2005) U.K. analyses suggest that education reform that delayed tracking produced a substantial increase in intergenerational mobility there, primarily to the benefit of children from low-income families, which cannot be ascribed to an increase in welfare state redistribution because over the period in question income inequality actually grew. More broadly, though, it has become increasingly clear that generalized policies promoting the attainment of higher levels of education by increasing the proportion going on to third level—assigned a central role in strategies aimed at improving equality of opportunity in many countries—may not be adequate if the aim is to address the disadvantages that children from poorer backgrounds face from the outset.

This has served to reinforce the emphasis in recent literature arguing for an early childhood focus, and that high-quality early childhood programs can significantly improve both cognitive and noncognitive outcomes for disadvantaged children (Carneiro and Heckman, 2003; Currie, 2001; Karoly et al., 2005; Waldfogel, 2006). Heckman's work has been particularly influential in demonstrating that investing in early childhood is a cost-effective policy (though the broader implications in terms of later interventions have been hotly debated). The core evidence that underpins Heckman's work comes from early intervention programs in the United States, but Esping-Andersen (2004a,b) relates the significant decline in social inheritance effects for the Nordic countries to the introduction of universal, high-quality child care. Schütz et al. (2005) in their cross-sectional comparison across countries report an inverted U-shaped relationship between family background effect and preschool enrolment, which suggests that early education may reduce the extent to which family background shapes life chances. OECD (2009) concludes that good quality care in early childhood, preschool, and also school years, are essential tools for promoting intergenerational mobility.

Going beyond education, the extent and nature of the welfare state itself can clearly affect the intergenerational transmission of poverty, indeed this is often articulated as a core aim in terms of equalizing children's life chances and avoiding wasted potential. One might expect that social policies that reduce child poverty (such as effective income support and promoting maternal employment, as discussed earlier—see UNICEF, 2007; Whiteford and Adema, 2007) would also promote more intergenerational inequality, but directly demonstrating that link is less straightforward. Mayer (1997), for example, argued that low income in itself is less important than parental characteristics such as low skills, poor health, or deviance, which affect the likelihood of being poor. In a comparison across the United States, though, Mayer and Lopoo (2008) find that in high-spending states the difference in mobility between advantaged and disadvantaged children is smaller than in low-spending states. It has been calculated that the risk of child poverty falls by a factor of four when mothers are employed (Esping-Andersen, 2009). There is also some evidence that intergenerational transmission of welfare dependency may be related to program design, with Corak et al.'s (2004) comparison of cash support schemes in North America and Sweden suggesting that passive programs are more likely to promote the transmission of welfare dependency than active ones. More generally, benefit systems that rely heavily on means-testing are more likely to create the poverty and unemployment traps that make it more likely that poverty and welfare dependency persist into subsequent generations.

Finally, still focusing on children and the transmission of poverty, an issue that has received considerable attention in the research literature is the potential effect of living in a "bad" neighborhood. Some studies suggest that local conditions can help explain the intergenerational transmission of income (OECD, 2008), although their impact may be relatively weak even in the United States. The range of U.S.-focused studies reported in Brooks-Gunn et al. (1997) suggested that neighborhood does matter for child and youth development, having greatest impact in early childhood and late adolescence and less in between, but the size of these effects was usually much smaller than those of family-level conditions. Solon et al. (2000) used the cluster sampling design of the Panel Study of Income Dynamics to estimate both sibling and neighborhood correlations on years of schooling and found sibling correlations of around 0.5, whereas their neighborhood estimates were as low as 0.1. Raaum et al. (2003) used Norwegian census data and concluded likewise that neighborhood correlations were small compared to sibling correlations, both for educational attainment and long-run earnings. This is consistent with the findings of U.S. experiments where families living in public housing were assigned housing vouchers by lottery encouraging them to move to neighborhoods with lower poverty rates; the results reported in Sanbonmatsu et al. (2006) show no significant effects on test scores. Looking beyond educational attainment to a broader set of poverty-related outcomes, the difficulties in adequately characterizing neighborhoods in terms of all their potentially relevant characteristics, and of distinguishing their effects on poverty and related outcomes from those of individual/family characteristics—taking into account that there may be interactions between them—have also been emphasized in research outside the United States (see Lupton, 2003).

23.4. THE WELFARE STATE, ANTIPOVERTY POLICY, AND THE ECONOMIC CRISIS OF THE LATE 2000s

23.4.1 Poverty, Income Inequality, and the Economic Crisis

The economic crisis experienced by the OECD countries since 2007–2008 has been the most serious since the Great Depression of the 1930s in terms of its impact on output and growth and is central to the ways in which poverty and antipoverty policies are now being thought about, studied, and debated. The crisis has affected poverty directly, as we will discuss, but it also has altered the context in which welfare states are currently operating and perspectives on how they are and should be evolving in the medium term. Here, we look first at the evidence on the immediate impact of the crisis, and then at the medium-term context for antipoverty policy.

The immediate impact of the crisis on income inequality and poverty has been the subject of a number of national and comparative studies, including Matsaganis and Leventi (2013), Callan et al. (2011), Figari et al. (2011), and Jenkins et al. (2013). Jenkins and colleagues adopt a comparative perspective, looking at aggregate indicators across the OECD and at six case-study countries in depth. Their central conclusion is that the immediate impact of the crisis on income inequality and income poverty in most countries was much more modest than the dramatic experience of the Great Depression, although not so different from more recent recessions, such as the Nordic crisis of the early 1990s. They stress that a striking feature of the crisis from 2007 to 2008 has been the extent to which its macroeconomic impact varied across countries: in some there were major declines in economic activity and sharply rising unemployment, but in others there was much more modest changes in growth and employment (see also Lane and Milesi-Ferretti, 2012). The peak-to-trough fall in quarterly GDP was substantially larger than the average fall during recessions over the previous 50 years almost everywhere but ranged, nonetheless, from zero in Australia to nearly 13% in Ireland. Another feature highlighted is that GDP declines were not fully transmitted into falls in the real disposable income of households, which were protected by both automatic stabilizers and additional support of governments through the tax and benefit system. The immediate response of employment to the fall in GDP was also frequently smaller than in previous recessions, though this was not the case in countries such as Ireland, Spain, and the United States where a boom–bust pattern in the housing market played an important role in the recession. Large falls in individual employment were also accompanied by significant rises in household worklessness in countries such as Ireland, Spain, and the United States, but not in some others—notably, Denmark and Finland where the workless household rate fell despite relatively large increases in the individual nonemployment rate, cushioning the impact on poverty. Another feature of the immediate onset of the crisis was the decline in income from capital, concentrated among richer households.

Looking at available poverty indicators up to 2009 compared with pre-crisis, Jenkins and colleagues found that relative income poverty rates typically fell in European countries, whereas absolute poverty rates (using "anchored" income thresholds indexed to prices) tended to fall slightly in Europe while rising modestly in the United States (as measured with the U.S. official poverty line), but in both cases these rates fell for the elderly. The six countries they studied in detail—Germany, Ireland, Italy, Sweden, the United Kingdom, and the United States—experienced differing macroeconomic shocks, with Germany recovering very rapidly, Sweden seeing a large decline in GDP but relatively rapid recovery, the United States experiencing marked contraction followed by some recovery, Italy and the United Kingdom seeing major downturns, and Ireland experiencing the largest GDP decline among OECD countries. Germany saw little change in employment, whereas in Ireland and in the United States at the other extreme, unemployment rose rapidly. The short-run impact on household income inequality and poverty was relatively modest. In Germany, the proportion of persons with a household income less than 60% of the contemporary median income declined marginally, and the proportion in households below such an income threshold held fixed in purchasing power at its 2007 level also fell. Chapter 2 shows that median income, inequality, and relative poverty all rose slightly in 2010. In the United Kingdom, the number falling below 60% of median income fell by more than 1 percentage point, and a fixed real threshold showed a larger decline in poverty. In Sweden, the proportion falling below 60% of median income increased, although when a threshold fixed in purchasing power terms is employed the increase was a good deal smaller. In Ireland, relative income poverty declined between 2007 and 2009 while the proportion below a fixed real income threshold remained stable. In Italy, the buffering role of social transfers was relatively limited, although the consequent increase in poverty might be considered modest given the scale of the initial macroeconomic shock. Finally, in the United States, the relative poverty rate declined modestly, reflecting a decline in real median income, whereas the official poverty rate (calculated using a low-income cut-off held fixed in real terms) increased. In all six case-study countries, elderly people were relatively well protected, compared with children and individuals of working age.

The variation in the distributional impact of the crisis to date across countries reflects not only differences in the nature of the macroeconomic downturn but also differences in how cash transfers and direct taxes cushioned household net incomes from the full effects of what was happening to market incomes. To some extent, these are differences in automatic stabilization and so they vary with the generosity and comprehensiveness of social safety nets and the structure and levels of direct taxes and social insurance contributions. However, policy responses and choices as the recession impacted have also been important (for a discussion of EU government's initial responses to the crisis see Marchal et al., 2014).

More recent poverty indicators for European countries produced by Eurostat, up to 2011, also show that experiences have been quite varied. As shown in Table 23.5, between 2007 and 2011 the proportion falling below 60% of median income rose by 1 percentage point or more in eight countries, fell by that amount in seven countries, and was stable in the rest. The average relative income poverty rate across the EU 27 was 16.5% in 2007 and 16.9% in 2011. Income poverty rates "anchored" at the 2008 60% of median threshold and then indexed to prices showed a good deal more variability over time across EU countries, as Table 23.6 shows. This rose in 13 countries, sometimes by a remarkably large amount—by 11 percentage points in Latvia and Lithuania and almost 14 percentage points in Iceland; however, it fell in another 10 countries, so that the overall average across the EU rose only from 16.4% to 17.5%. It is interesting to compare this with the trend in material deprivation over the same period, as measured by the EU's severe material deprivation indicator: Table 23.7 shows that this rose between 2008 and 2011 in 13 countries while falling in 6; the average across the EU rose marginally. Among countries particularly hard hit by the crisis, deprivation rose sharply in Ireland, Spain, Greece, and Italy, as well as in Latvia and Lithuania, but fell in Portugal.

23.4.2 The Crisis and Antipoverty Policy in the Medium Term

The immediate impact of the onset of the crisis from 2007 to 2008 on living standards and poverty was cushioned, at least to some extent, by welfare state institutions and in particular by social protection and tax systems. The medium-term impact of the crisis on poverty depends not only on developments in the macroeconomy and in employment, but also on the policies adopted with respect to the welfare state broadly conceived and to transfers most particularly. The effects of the crisis on the public finances are dominant in framing the context in which these choices are being made. The need—or perception of such a need—to consolidate public finances plays a central role in debates about responding to the crisis, with tackling poverty often relegated to a more modest role. This could lead to changes to welfare state systems and parameters that will take many years to work their way through, continuing to have an impact on poverty long after economic growth has resumed and the recession is considered to have ended from a purely macroeconomic perspective. (The fairness of fiscal consolidation programs may itself affect the likelihood of them being successful, as analyzed by Kaplanoglou et al. (2013) for 29 OECD countries over the period 1971–2009; their results suggest that programs improving the targeting of

Table 23.5 Relative income poverty rates (60% of median threshold), European Union countries 2007–2011

Country	2007 (%)	2008 (%)	2009 (%)	2010 (%)	2011 (%)
Belgium	15.2	14.7	14.6	14.6	15.3
Bulgaria	22.0	21.4	21.8	20.7	22.3
Czech Republic	9.6	9.0	8.6	9.0	9.8
Denmark	11.7	11.8	13.1	13.3	13.0
Germany	15.2	15.2	15.5	15.6	15.8
Estonia	19.4	19.5	19.7	15.8	17.5
Ireland	17.2	15.5	15.0	16.1	–
Greece	20.3	20.1	19.7	20.1	21.4
Spain	19.7	19.6	19.5	20.7	21.8
France	13.1	12.7	12.9	13.3	14.0
Italy	19.8	18.7	18.4	18.2	19.6
Cyprus	15.5	15.9	15.8	15.1	14.5
Latvia	21.2	25.6	25.7	21.3	19.1
Lithuania	19.1	20.0	20.6	20.2	20.0
Luxembourg	13.5	13.4	14.9	14.5	13.6
Hungary	12.3	12.4	12.4	12.3	13.8
Malta	14.8	15.0	15.3	15.0	15.4
Netherlands	10.2	10.5	11.1	10.3	11.0
Austria	12.0	12.4	12.0	12.1	12.6
Poland	17.3	16.9	17.1	17.6	17.7
Portugal	18.1	18.5	17.9	17.9	18.0
Romania	24.8	23.4	22.4	21.1	22.2
Slovenia	11.5	12.3	11.3	12.7	13.6
Slovakia	10.6	10.9	11.0	12.0	13.0
Finland	13.0	13.6	13.8	13.1	13.7
Sweden	10.5	12.2	13.3	12.9	14.0
United Kingdom	18.6	18.7	17.3	17.1	16.2
Iceland	10.1	10.1	10.2	9.8	9.2
Norway	11.9	11.4	11.7	11.2	10.5
Switzerland	–	16.2	15.1	15.0	15.0
Croatia	18	17.3	17.9	20.5	21.1
European Union (27 countries)	16.5	16.4	16.3	16.4	16.9

Notes: The household income statistics in Eurostat are mainly produced with EU-SILC data, which reference period is a fixed 12-month period (such as the previous calendar or tax year) for all countries except the United Kingdom for which the income reference period is the current year and IE for which the survey is continuous and income is collected for the last 12 months.
Source: Eurostat (downloaded March 20, 2013).

social transfers and their effectiveness in poverty alleviation, increasing spending on training and active labor market policies, and even reducing value-added taxes on necessities, enhance the probability of successful adjustment while promoting social cohesion.)

In such a context, the pressure to increase the targeting of cash transfers is likely to intensify, although that can run the risk of worsening poverty and unemployment

Table 23.6 Anchored income poverty rates (60% of median threshold in 2008, indexed to consumer prices subsequently), European Union countries 2008–2011

Country	2008 (%)	2009 (%)	2010 (%)	2011 (%)
Belgium	14.7	13.1	13.0	13.5
Bulgaria	21.4	16.1	14.8	17.8
Czech Republic	9.0	8.1	7.8	8.6
Denmark	11.8	13.1	12.6	12.2
Germany	15.2	16.0	15.8	15.9
Estonia	19.5	18.9	19.7	23.9
Ireland	15.5	15.4	22.8	–
Greece	20.1	18.9	18.0	24.9
Spain	19.6	20.2	22.3	25.7
France	12.7	12.7	12.3	13.9
Italy	18.7	19.9	19.3	21.4
Cyprus	15.9	16.3	16.2	14.4
Latvia	25.6	26.0	33.0	36.2
Lithuania	20.0	18.6	28.4	30.8
Luxembourg	13.4	15.5	14.4	14.6
Hungary	12.4	11.8	13.7	14.7
Malta	15.0	14.3	16.5	15.9
Netherlands	10.5	10.6	10.0	11.0
Austria	12.4	11.4	11.0	10.5
Poland	16.9	13.7	13.0	11.9
Portugal	18.5	18.1	16.1	17.9
Romania	23.4	18.2	16.2	17.9
Slovenia	12.3	10.2	12.1	13.0
Slovakia	10.9	7.8	7.3	7.0
Finland	13.6	13.0	12.0	12.3
Sweden	12.2	11.7	11.2	11.6
United Kingdom	18.7	20.4	21.4	21.8
Iceland	10.1	9.8	16.7	23.7
Norway	11.4	10.2	9.6	8.9
Switzerland	16.2	13.8	13.8	13.1
European Union (27 countries)	16.4	16.3	16.4	17.5

Notes: The household income statistics in Eurostat are mainly produced with EU-SILC data, which reference period is a fixed 12-month period (such as the previous calendar or tax year) for all countries except the United Kingdom for which the income reference period is the current year and IE for which the survey is continuous and income is collected for the last 12 months.
Source: Eurostat (downloaded March 20, 2013).

"traps" and undermining the bases for social solidarity and political support for relatively generous provision. The notion of "social investment" has come to play a major part in debates about the role of social spending and the future of welfare states in rich countries, particularly in Europe where the language of social investment has become embedded in EU discourse since the adoption of the Lisbon Agenda in 2000. A number of important

Table 23.7 Severe material deprivation rate, European Union countries 2008–2011

Country	2007 (%)	2008 (%)	2009 (%)	2010 (%)	2011 (%)
Belgium	5.7	5.6	5.2	5.9	5.7
Bulgaria	57.6	41.2	41.9	45.7	43.6
Czech Republic	7.4	6.8	6.1	6.2	6.1
Denmark	3.3	2.0	2.3	2.7	2.6
Germany	4.8	5.5	5.4	4.5	5.3
Estonia	5.6	4.9	6.2	9.0	8.7
Ireland	4.5	5.5	6.1	7.5	–
Greece	11.5	11.2	11.0	11.6	15.2
Spain	3.0	2.5	3.5	4.0	3.9
France	4.7	5.4	5.6	5.8	5.2
Italy	6.8	7.5	7.0	6.9	11.2
Cyprus	13.3	9.1	9.5	10.1	10.8
Latvia	24.9	19.0	21.9	27.4	31.4
Lithuania	16.6	12.3	15.1	19.5	18.5
Luxembourg	0.8	0.7	1.1	0.5	1.2
Hungary	19.9	17.9	20.3	21.6	23.1
Malta	4.2	4.0	4.7	5.7	6.3
Netherlands	1.7	1.5	1.4	2.2	2.5
Austria	3.3	6.4	4.8	4.3	3.9
Poland	22.3	17.7	15.0	14.2	13.0
Portugal	9.6	9.7	9.1	9.0	8.3
Romania	36.5	32.9	32.2	31.0	29.4
Slovenia	5.1	6.7	6.1	5.9	6.1
Slovakia	13.7	11.8	11.1	11.4	10.6
Finland	3.6	3.5	2.8	2.8	3.2
Sweden	2.2	1.4	1.6	1.3	1.2
United Kingdom	4.2	4.5	3.3	4.8	5.1
Iceland	2.1	0.8	0.8	1.8	2.1
Norway	2.3	2.0	2.2	2.0	2.3
Switzerland	–	2.2	2.1	1.7	1.0
Croatia	–	–	–	14.5	14.8
European Union (27 countries)	9.1	8.4	8.1	8.3	8.8

Source: Eurostat (downloaded March 20, 2013).

recent contributions have highlighted the potential of social investment as a new perspective on or a paradigm for social policy in the context of the economic crisis and to the demand of the knowledge-based economy more broadly, as an alternative to neoliberal responses focusing on retrenchment in social spending, and as a key ingredient in responding to the macroeconomic/Euro crisis (see the contributions to Hemerijck and Vandenbroucke, 2012; Morel et al., 2011; Vandenbroucke et al., 2011). Others have sought to assess the extent to which recent directions in social policies and spending patterns could be characterized as moving toward a social investment strategy and whether disappointing outcomes in terms of poverty can be seen as a failure of such a strategy (Cantillon, 2011; Vandenbroucke and Vleminckx, 2011; van Kersbergen and Hemerijck, 2012). The EU is paying serious attention to this debate, as evidenced by the establishment by DG Employment, Social Affairs and Equal Opportunities of an expert group on Social Investment for Growth and Cohesion in autumn 2012 as input to a major initiative envisaged in the area of social policies.

"Social investment" may be viewed in a number of distinct ways, as Nolan (2013) discusses: as a paradigm and strategy for social policies and spending, as a conceptual base and analytical framework, and/or as a platform for political engagement in both a narrow and broad sense. Whether social investment can credibly be presented as the paradigm most likely to underpin economic growth or employment is open to debate and merits further research, even if—as Nolan (2013) argues—the distinction between social "investment" and other social spending is not particularly robust, conceptually and empirically. Highlighting that distinction may not in any case be the most useful and productive way to frame the debate about the future of social spending, where concentration on a narrow economic argument runs the risk of obscuring normative choices and the broader case for social spending.

Finally, it is important to note that an economic crisis of the depth and nature of the one that began in 2007–2008 may also have major implications for intergenerational equity, especially if it continues to be the case that the elderly are relatively well-cushioned from its effects compared to younger people; sustained high unemployment in particular may well result in long-term "scarring" of those affected, with the risk that their disadvantage is transmitted to the next generation.

23.5. FUTURE RESEARCH DIRECTIONS

We conclude with a brief discussion of priorities for research on poverty and antipoverty policy. The key challenges lie in deepening understanding of the processes at work in creating and perpetuating poverty at individual, household, national, and cross-national level. While much has been learned about the characteristics associated with poverty in different countries, the fact that this differs so widely across countries provides a window into the nature of the underlying processes that has not been fully exploited. In the same vein, studying the factors associated with change over time in a specific country is valuable but putting these changes in a comparative perspective adds another dimension. So a panel-of-countries approach has increasing potential as the statistical underpinning in terms of comparable data continues to be built. This can be complemented by continued development of the potential to carry out micro-simulation analysis in a comparative perspective; the challenge of incorporating behavioral responses into such analysis remains substantial (Immervoll et al., 2007). Exploiting the potential of panel data will continue to be a priority to reliably distinguish those data genuinely and persistently on low income and understanding the barriers to income smoothing facing those on low income more transiently. Increasing recognition of the multidimensional nature of poverty and social exclusion points to the need to deepen understanding of the linkages between different forms of deprivation and exclusion, moving beyond descriptive analysis of the extent to which they go together to study the processes that underpin the underlying relationships between them—where once again a comparative perspective is invaluable—while also addressing the difficult conceptual issues involved.

There also remains a substantial research agenda in the field of antipoverty policy. Not many countries have made very substantial progress in reducing relative poverty as conventionally measured in recent years, though material deprivation and absolute poverty have generally declined up to the crisis from 2008. While some progress has been made in understanding the factors at work, many of the deeper causal questions remain largely unsettled. Changes in the distribution of income from the market may have made reducing relative poverty more difficult, and the redistributive impact of tax and benefit systems may have declined, and each needs to be much better understood. A key question is whether the apparent failure of many governments to maintain or to improve the antipoverty impact of their tax and benefit systems is a consequence of lack of effective political will (voter preferences) or reflects instead (or as well) systemic limits and/or external constraints. Important items on the policy research agenda include:

- Can more be done with less? There is a continuing controversy over targeting and cost-effectiveness of public social expenditure. With ageing populations and rising needs due to sociodemographic and economic trends, this question is bound to remain at the forefront of the research agenda.
- Why are antipoverty provisions in many countries so manifestly inadequate? Are there systemic limits to incrementalism in redistributive policy? That is to say: are there really limits to what improvement can be achieved by strengthening the existing main pillars of redistribution: wage and broader market force regulation, social insurance, social assistance, and taxes? What promise do new redistributive mechanisms and programs offer? Negative income taxes and associated systems are seen as the way forward by some, but short-term issues, such as take-up and long-term effects on wages and human capital formation, earnings mobility, and so forth are not well understood.
- What is the optimal balance between direct redistribution and "social investment"—that is, expenditures that seek to generate lasting effects through improvements in skills and capabilities? To what extent can social investment act as a substitute for direct "compensatory" redistribution, or is there complementarity? If so, what is the optimal balance?
- Making cash benefits and services conditional on certain behavioral requirements and conditions is a policy strategy that is gaining increased attention, part of a broader

current toward more micro intervention in social policy, and informed by social experiments (see Bastagli, 2011; Medgyesi and Temesváry, 2013). Is such a shift from the macro to the micro level really the way forward, and what, if any, are the limits there?

Finally, we should note that while this survey has focused on the "rich world" (as it is conventionally understood), some of the most innovative antipoverty policy is being conceived, implemented, and analyzed outside of that area, with a number of South American and Asian countries standing out in this respect. An important task for future research is to integrate these rich but largely parallel streams of poverty research.

ACKNOWLEDGMENTS

We thank the participants at the April 2013 Conference "Recent Advances in the Economics of Income Distribution" held at the Paris School of Economics and organized by A.B. Atkinson and F. Bourguignon. Particularly, thanks go to Rolf Aaberge, who served as the main discussant of this chapter.

REFERENCES

Aaberge, R., Flood, L., 2013. U.S. versus Sweden: the effect of alternative in-work tax credit policies on labour supply of single mothers: IZA Discussion Papers 7706. Institute for the Study of Labor (IZA), Bonn.

Aaberge, R., Langørgen, A., 2006. Measuring the benefits from public services: the effects of local government spending on the distribution of income in Norway. Rev. Income Wealth 52 (1), 61–83.

Aaberge, R., Mogstad, M., 2012. Inequality in current and lifetime income: Discussion Papers 726. Research Department of Statistics Norway.

Aaberge, R., Peluso, E., 2012. A counting approach for measuring multidimensional deprivation: IZA Discussion Papers 6589. Institute for the Study of Labor (IZA), Bonn.

Aaberge, R., Björklund, A., Jäntti, M., Palme, M., Pedersen, P., Smith, N., Wennemo, T., 2002. Income inequality and income mobility in the Scandinavian countries compared to the United States. Rev. Income Wealth 48 (4), 443–469.

Aaberge, R., Bhuller, M., Langørgen, A., Mogstad, M., 2010. The distributional impact of public services when needs differ. J. Public Econ. 94 (9–10), 549–562.

Adema, W., Fron, P., Ladaique, M., 2011. Is the European Welfare State Really More Expensive? Indicators on Social Spending, 1980–2012; and a Manual to the OECD Social Expenditure Database (SOCX). OECD Publishing, Paris.

Airio, I., 2008. Change of Norm? In-Work Poverty in a Comparative Perspective: Studies in Social Security and Health 92. Kela Research Department, KELA, Helsinki.

Allègre, G., Jaehrling, K., 2011. Making work pay for whom? Tax and benefits impacts on in-work poverty. In: Fraser, N., Gutiérrez, R., Peña-Casas, R. (Eds.), Working Poverty in Europe: A Comparative Approach. Palgrave Macmillan, Basingstoke, pp. 278–303.

Andreß, H.-J., Lohmann, H., 2008. The Working Poor in Europe: Employment, Poverty and Globalization. Edward Elgar, Cheltenham.

Arza, C., Kohli, M., 2008. Changing European Welfare: The New Distributional Principles of Pension Policy. Routledge, New York.

Ashenfelter, O., Card, D., 1999. first ed. Handbook of Labor Economics, vol. 3. Elsevier/North Holland, Amsterdam.

Ashenfelter, O., Card, D., 2011. first ed. Handbook of Labor Economics, vol. 4. Elsevier/North Holland, Amsterdam.

Ashenfelter, O., Layard, R., 1987. first ed. Handbook of Labor Economics, vol. 2. Elsevier/North Holland, Amsterdam.
Atkinson, A.B., 1987. On the measurement of poverty. Econometrica 55 (4), 749–764.
Atkinson, A.B., 2003. Multidimensional deprivation: contrasting social welfare and counting approaches. J. Econ. Inequal. 1 (1), 51–65.
Atkinson, A.B., Marlier, E., 2010. Income and Living Conditions in Europe. Eurostat, Luxembourg.
Atkinson, A., Rainwater, L., Smeeding, T., 1995. Income Distribution in OECD Countries. OECD, Paris.
Atkinson, T., Cantillon, B., Marlier, E., Nolan, B., 2002. Social Indicators: The EU and Social Inclusion. Oxford University Press, Oxford.
Bane, M.J., Ellwood, D., 1986. Slipping in and out of poverty: the dynamics of poverty spells. J. Hum. Resour. 12, 1–23.
Barbier, J.-C., Ludwig-Mayerhofer, W., 2004. Introduction: the many worlds of activation. Eur. Soc. 6 (4), 423–436.
Bargain, O., Orsini, K., 2006. In-work policies in Europe: killing two birds with one stone? Labour Econ. 13 (6), 667–697.
Bargain, O., Orsini, K., 2007. Beans for breakfast? How exportable is the British workfare model? In: Bargain, O. (Ed.), Microsimulation in Action. Policy Analysis in Europe using EUROMOD. Research in Labour Economics, vol. 25. Elsevier, Oxford.
Barr, N., 2001. The Welfare State as Piggy Bank: Information, Risk, Uncertainty and the Role of the State. Oxford University Press, Oxford.
Barth, E., Moene, K., 2009. The equality multiplier: NBER Working Paper 15076. National Bureau of Economic Research, Cambridge, MA.
Bastagli, F., 2011. Conditional cash transfers as a tool of social policy. Econ. Polit. Wkly XLVI, 61–66.
Becker, G., 1964. Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education. University of Chicago Press, Chicago.
Bekker, S., Wilthagen, A.C.J.M., 2008. Flexicurity: a European approach to labour market policy. Intereconomics 43 (2), 68–73.
Benedict, M.E., Shaw, K., 1995. The impact of pension benefits on the distribution of earned income. Ind. Labor Relat. Rev. 48 (4), 740–757.
Beramendi Alvarez, P., 2001. The politics of income inequality in the OECD. The role of second order effects: Luxembourg Income Study Working Paper No. 284.
Bergh, A., 2005. On the counterfactual problem of welfare state research: how can we measure redistribution? Eur. Soc. Rev. 21 (4), 345–357.
Biblarz, T., Raftery, A., 1999. Family structure, educational attainment, and socioeconomic success: rethinking the "pathology of matriarchy" Am. J. Soc. 105, 321–365.
Blanden, J., Gibbons, S., 2006. The Persistence of Poverty Across Generations: A View from Two Cohorts. Policy Press, Bristol.
Blanden, J., Gregg, P., Machin, S., 2005. Educational inequality and intergenerational mobility. In: Machin, S., Vignoles, A. (Eds.), What's the Good of Education? Princeton University Press, Princeton, NJ.
Blau, F., Kahn, L., 2008. Inequality and earnings distribution. In: Salverda, W., Nolan, B., Smeeding, T. (Eds.), The Oxford Handbook of Economic Inequality. Oxford University Press, Oxford.
Boarini, R., Mira d'Ercole, M., 2006. Measures of Material Deprivation in OECD Countries: OECD Social, Employment and Migration Working Papers 2006(6). OECD, Paris.
Bonoli, G., 2007. Time matters. Postindustrialisation, new social risks and welfare state adaptation in advanced industrial democracies. Comp. Polit. Stud. 40, 495–520.
Bonoli, G., 2010. The political economy of active labour market policies. Polit. Soc. 38, 435–457.
Börsch-Supan, A.H., 2012. Entitlement reforms in Europe: policy mixes in the current pension reform process: NBER Working Paper No. 18009.
Bourguignon, F., Chakravarty, S., 2003. The measurement of multidimensional poverty. J. Econ. Inequal. 1 (1), 25–49.
Bourguignon, F., Rogers, F.H., 2007. Distributional effects of educational improvements: are we using the wrong model? Econ. Educ. Rev. 26 (6), 735–746.

Bovenberg, L., van Ewijk, C., 2011. The future of multi-pillar pension systems: Netspar Discussion Papers DP 09/2011-079.
Bradshaw, J., 2010. An international perspective on child benefit packages. In: Kamerman, S., Phipps, S., Ben-Arieh, A. (Eds.), From Child Welfare to Child Well-Being: An International Perspective on Knowledge in the Service of Policy Making. Springer, Berlin, pp. 293–307.
Bradshaw, J., Finch, N., 2002. A comparison of child benefit packages in 22 countries: Research Report 174. Department for Work and Pensions, Norwich.
Bradshaw, J., et al., 2008. A Minimum Income Standard for Britain. What People Think. Joseph Rowntree Foundation, York.
Brooks-Gunn, J., Duncan, G.J., Aber, J.L., 1997. Neighborhood Poverty. Context and Consequences for Children, vol. 1 Russell Sage Foundation, New York, NY.
Bruno, R., Rovelli, R., 2010. Labour market policies and outcomes in the enlarged EU. J. Common Market Stud. 48, 661–685.
Buhmann, B., Rainwater, L., Schmaus, G., Smeeding, T., 1987. Equivalence scales, well-being, inequality and poverty: sensitivity estimates across ten countries using the Luxembourg income study (LIS) database. Rev. Income Wealth 34, 115–142.
Burniaux, J.-M., Mira d'Ercole, M., 2006. Labour market performance, income inequality and poverty in OECD countries: OECD Economics Department Working Paper 500. OECD, Paris.
Burniaux, J., Padrini, F., Brandt, N., 2006. Labour Market Performance, Income Inequality and Poverty in OECD Countries. OECD Economics Department Working Papers, No. 500, OECD Publishing.
Burniaux, J.-M., Dan, T.-T., Fore, D., Förster, M., Mira d'Ercole, M., Oxley, H., 1998. Income Distribution and Poverty in Selected OECD Countries: OECD Economics Department Working Paper 189. OECD, Paris.
Burtless, G., 2006. Poverty and inequality. In: Clark, G.L., Munell, A.H., Orszag, J.M. (Eds.), The Oxford Handbook of Pensions and Retirement Income. Oxford University Press, Oxford.
Burtless, G., 2009. Demographic transformation and economic inequality. In: Salverda, W., Nolan, B., Smeeding, T.M. (Eds.), The Oxford Handbook of Economic Inequality. Oxford University Press, Oxford.
Callan, T., Keane, C., 2009. Non-cash benefits and the distribution of economic welfare. Econ. Soc. Rev. 40 (1), 49–71.
Callan, T., Smeeding, T., Tsakloglou, P., 2008. Short-run distributional effects of public education transfers to tertiary education students in seven European countries. Educ. Econ. 16 (3), 275–288.
Callan, T., Nolan, B., Walsh, J., 2011. The economic crisis, public sector pay, and the income distribution. In: Immervoll, H., Peichl, A., Tatsiramos, K. (Eds.), In: Research on Labor Economics, vol. 32. Emerald, Bingley, pp. 207–225.
Caminada, K., Goudswaard, K., Wang, C., 2012. Disentangling income inequality and the redistributive effect of taxes and transfers in 20 LIS countries over time: LIS Working Paper Series Luxembourg Income Study (LIS) No. 581.
Cantillon, B., 2011. The paradox of the social investment state: growth, employment and poverty in the Lisbon era. J. Eur. Soc. Policy 21 (5), 432–449.
Cantillon, B., Marx, I., Van den Bosch, K., 2003. The puzzle of egalitarianism: the relationship between employment, wage inequality, social expenditure and poverty. Eur. J. Soc. Secur. 5 (2), 108–127.
Cantillon, B., Van Mechelen, N., Pintelon, O., van den Heede, A., 2014. Social redistribution, poverty and the adequacy of social protection in the EU. In: Cantillon, B., Vandenbroucke, F. (Eds.), Reconciling Work and Poverty Reduction: How Successful Are European Welfare States? Oxford University Press, Oxford, pp. 157–184.
Card, D., 1999. The causal effect of education on earnings. In: Ashenfelter, O., Card, D. (Eds.), In: Handbook of Labor Economics, vol. 3. Elsevier/North Holland, Amsterdam.
Card, D., Kluve, J., Weber, A., 2010. Active labor market policy evaluations: a meta-analysis. Econ. J. 120, F452–F477.
Card, D., Heining, J., Kline, P., 2013. Workplace heterogeneity and the rise of West German wage inequality. Q. J. Econ. 128 (3), 967–1015.

Carneiro, P., Heckman, J., 2003. Human capital policy. In: Heckman, J., Krueger, A. (Eds.), Inequality in America. MIT Press, Cambridge, MA.
CERC, 2004. Child Poverty in France. Conseil de L'Emploi, des Revenues et de la Cohesion Sociale, Paris.
Checchi, D., van de Werfhorst, H., Braga, M., Meschi, E., 2014. The policy response: education. In: Salverda, W., Nolan, B., Checchi, D., Marx, I., McKnight, A., Tóth, I., van de Werfhorst, H. (Eds.), Changing Inequalities and Societal Impacts in Rich Countries: Analytical and Comparative Perspectives. Oxford University Press, Oxford.
Chetty, R., 2009. Sufficient statistics for welfare analysis: a bridge between structural and reduced-form methods. Ann. Rev. Econ. 1 (1), 451–488.
Chetty, R., Friedman, J.N., Saez, E., 2013. Using differences in knowledge across neighborhoods to uncover the impacts of the EITC on earnings: NBER Working Paper 18232.
Citro, C., Michael, R., 1995. Measuring Poverty: A New Approach. National Research Council, National Academy Press, Washington, DC.
Corak, M., 2001. Are the kids all right? Intergenerational mobility and child well-being in Canada: Analytical Studies Branch Research Paper Series 2001171e. Statistics Canada, Ottawa.
Corak, M., 2004. Generational income mobility in North America and Europe: an introduction. In: Corak, M. (Ed.), Generational Income Mobility in North America and Europe. Cambridge University Press, Cambridge.
Corak, M., Gustafsson, B., Österberg, T., 2004. Intergenerational influences on the receipt of unemployment insurance in Canada and Sweden. In: Corak, M. (Ed.), Generational Income Mobility in North America and Europe. Cambridge University Press, Cambridge.
Corak, M., Lietz, C., Sutherland, H., 2005. The impact of tax and transfer systems on children in the European Union: IZA Discussion Paper 1589. Institute for the Study of Labor, Bonn.
Corcoran, M., 2001. Mobility, persistence, and the consequences of poverty for children: child and adult outcomes. In: Danziger, S., Haveman, R. (Eds.), Understanding Poverty. Russell Sage Foundation/Harvard University Press, Cambridge, MA.
Corluy, V., Vandenbroucke, F., 2013. Household joblessness. In: Cantillon, B., Vandenbroucke, F. (Eds.), Reconciling Work and Poverty Reduction. How Successful Are European Welfare States? Oxford University Press, Oxford.
Coulter, F., Cowell, F., Jenkins, S., 1992. Equivalence scale relativities and the extent of inequality and poverty. Econ. J. 102, 1067–1082.
Crettaz, E., 2011. Fighting Working Poverty in Post-Industrial Economies. Causes, Trade-Offs and Policy Solutions. Edward Elgar, Cheltenham, UK/Northampton, MA, USA.
Cunha, F., Heckman, J., 2007. The technology of skill formation. Am. Econ. Rev. 97 (2), 31–47.
Currie, J., 2001. Early childhood intervention programs. J. Econ. Perspect. 15 (2), 213–238.
Currie, J., Gahvari, F., 2008. Transfers in cash and in-kind: theory meets the data. J. Econ. Lit. 46 (2), 333–383.
D'Addio, A.C., 2007. Intergenerational transmission of disadvantage: mobility or immobility across generations? A review of the evidence for OECD countries: Social, Employment and Migration Working Papers No. 52. OECD, Paris.
Dang, T.-T., Immervoll, H., Mantovani, D., Orsini, K., Sutherland, H., Dang, T.-T., Immervoll, H., Mantovani, D., Orsini, K., Sutherland, H., 2006. An age perspective on economic well-being and social protection in nine OECD countries: IZA Discussion Papers 2173. Institute for the Study of Labor (IZA), Bonn.
Danziger, S., Waldfogel, J., 2000. Securing the Future: Investing in Children From Birth to College. Russell Sage Foundation, New York.
De Deken, J., Kittel, B., 2007. Social expenditure under scrutiny: the problems of using aggregate spending data for assessing welfare state dynamics. In: Clasen, J., Siegel, N.A. (Hg.), Investigating Welfare State Change: The "Dependent Variable Problem" in Comparative Analysis. Edward Elgar Publishing, Cheltenham, pp. 72–104.
De Graaf, N.D., de Graaf, P., Kraaykamp, G., 2000. Parental cultural capital and educational attainment in The Netherlands: a refinement of the cultural capital perspective. Soc. Educ. 73 (2), 92–111.

De Graaf-Zijl, M., Nolan, B., 2011. Household joblessness and its impact on poverty and deprivation in Europe. J. Eur. Soc. Policy 21 (5), 413–431.
Deacon, A., Bradshaw, J., 1983. Reserved for the Poor: The Means Test in British Social Policy. Robertson, Oxford.
Deaton, A., 2010. Instruments, randomization, and learning about development. J. Econ. Lit. 48, 424–455.
Deaton, A., Gourinchas, P.-O., Paxson, C., 2002. Social security and inequality over the life cycle. In: Feldstein, M., Leibman, J. (Eds.), The Distributional Effects of Social Security Reform. Chicago University Press for NBER, Chicago, pp. 115–148.
Dickens, R., Ellwood, D.T., 2003. Child poverty in Britain and the United States. Econ. J. 113 (488), F219–F239.
Digeldey, I., 2007. Between workfare and enablement—the different paths to transformation of the welfare state: a comparative analysis of activating labour market policies. Eur. J. Polit. Res. 46 (6), 823–851.
Dolls, M., Fuest, C., Peichl, A., 2011. Automatic stabilizers, economic crisis and income distribution in Europe. In: Immervoll, H., Peichl, A., Tatsiramo, K. (Eds.), Who Loses in the Downturn? Economic Crisis, Employment and Income Distribution. In: Research in Labor Economics, vol. 32. Emerald Group Publishing Limited, Bingley, pp. 227–255.
Dolls, M., Fuest, C., Peichl, A., 2012. Automatic stabilizers and economic crisis: US vs. Europe. J. Pub. Econ. 96, 279–294.
Duclos, J.-Y., Makdissi, P., 2005. Sequential stochastic dominance and the robustness of poverty orderings. Rev. Income Wealth 51 (1), 63–87.
Duncan, G.J., Brooks-Gunn, J., 1997. Consequences of Growing Up Poor. Russell Sage Foundation, New York.
Duncan, G., Gustafsson, B., Hauser, R., Schmaus, G., Messinger, H., Muffels, R., Nolan, B., Ray, J.-C., 1993. Poverty dynamics in eight countries. J. Popul. Econ. 6 (3), 215–234.
Duncan, G.J., Brooks-Gunn, J., Klebanov, P.K., 1994. Economic deprivation and early childhood development. Child Dev. 65 (2), 296–318.
Duncan, G.J., Yeung, W.J., Brooks-Gunn, J., Smith, J.R., 1998. How much does childhood poverty affect the life chances of children? Am. Soc. Rev. 63, 406–423.
Eichhorst, W., Gienberger-Zingerle, M., Konle-Seidl, R., 2008. Activation policies in Germany: from status protection to basic income support. In: Eichhors, O., Kaufmann, O., Konle-Seidl, R. (Eds.), Bringing the Jobless into Work? Experiences with Activation Schemes in Europe and the US. Springer, Berlin.
Eissa, N., Hoynes, H., 2004. Taxes and the labor market participation of married couples: the earned income tax credit. J. Public Econ. 88 (9–10), 1931–1958.
Erikson, R., 1993. Descriptions in inequality: the Swedish approach to welfare research. In: Nussbaum, M.C., Sen, A. (Eds.), The Quality of Life. Clarendon Press, Oxford.
Esping-Andersen, G., 1990. The Three Worlds of Welfare Capitalism. Polity Press, Cambridge, UK.
Esping-Andersen, G., 1999. Social Foundations of Postindustrial Economies. Oxford University Press, Oxford.
Esping-Andersen, G., 2004a. Unequal opportunities and the mechanisms of social inheritance. In: Corak, M. (Ed.), Generational Income Mobility in North America and Europe. Cambridge University Press, Cambridge.
Esping-Andersen, G., 2004b. Untying the Gordian knot of social inheritance. Res. Soc. Stratif. Mobil. 21, 115–139.
Esping-Andersen, G., 2007. Sociological explanations of changing income distributions. Am. Behav. Sci. 50 (5), 639–658.
Esping-Andersen, G., 2009. The Incomplete Revolution: Adapting to Women's New Roles. Polity Press, Cambridge.
Esping-Andersen, G., Wagner, S., 2010. Asymmetries in the opportunity structure: intergenerational mobility trends in Scandinavia and Continental Europe: Department of Sociology Working Paper. UPF, Barcelona.
Esping-Andersen, G., Gallie, D., Hemerijck, A., Myles, J., 2002. Why We Need a New Welfare State. Oxford University Press, Oxford.

EU Commission, 2013. Communication from the Commission: Towards Social Investment for Growth and Cohesion – including implementing the European Social Fund 2014–2020. COM(2013) 83 final, European Commission, Brussels.
European Commission, 2008. Employment in Europe. Publications Office of the European Communities, Luxembourg.
European Union, 2013. Social Investment Package. European Union, Brussels.
Eurostat, 2005. Material Deprivation in the EU, Statistics in Focus, Population and Social Conditions, 21/2005. Statistical Office of the European Communities, Luxembourg.
Ferrarini, T., Nelson, K., 2003. Taxation of social insurance and redistribution: a comparative analysis of ten welfare states. J. Eur. Soc. Policy 13 (1), 21–33.
Ferrera, M., 1996. The 'southern model' of welfare in social Europe. J. Eur. Soc. Policy 6 (1), 17–37.
Figari, F., Sutherland, H., 2013. EUROMOD: the European Union tax-benefit microsimulation model. J. Microsimul. 6 (1), 4–26.
Figari, F., Salvatori, A., Sutherland, H., 2011. Economic downturn and stress testing European welfare systems. In: Immervoll, H., Peichl, A., Tatsiramos, K. (Eds.), Who Loses in the Downturn? Economic Crisis, Employment and Income Distribution. In: Research in Labor Economics, vol. 32. Emerald Group Publishing Limited, Bingley, pp. 257–286.
Förster, M., Mira d'Ercole, M., 2005. Income Distribution and Poverty in OECD Countries in the Second Half of the 1990s: OECD Social Employment and Migration Working Papers No. 22. OECD, Paris.
Förster, M., Pearson, M., 2002. Income distribution and poverty in the OECD area: trends and driving forces. OECD Econ. Stud. 34, 7–39.
Foster, J., Greer, W.J., Thorbecke, E., 1984. A class of decomposable poverty indices. Econometrica 52, 761–766.
Fouarge, D., Layte, R., 2005. Welfare regimes and poverty dynamics: the duration and recurrence of poverty spells in Europe. J. Soc. Policy 34, 1–20.
Fraser, N., Gutiérrez, R., Peña-Casas, R., 2011. Working Poverty in Europe: A Comparative Approach. Palgrave Macmillan, Basingstoke.
Freeman, R., 1998. War of the models: which labour market institutions for the 21st century? Labour Econ. 5, 1–24.
Freeman, R., 2008. Globalization and inequality. In: Salverda, W., Nolan, B., Smeeding, T. (Eds.), The Oxford Handbook of Economic Inequality. Oxford University Press, Oxford.
Fritzell, J., Ritakallio, V.-M., 2004. Societal shifts and changed patterns of poverty: Luxembourg Income Study Working Paper No. 393. LIS, Luxembourg.
Garfinkel, I., Rainwater, L., Smeeding, T., 2006. Wealth and Welfare States: Is America aLaggard or Leader? Oxford University Press, Oxford.
Gassmann, F., Notten, G., 2008. Size matters: poverty reduction effects of means-tested and universal child benefits in Russia. Eur. J. Soc. Policy 18 (3), 260–274.
Giannelli, G.C., Jaenichen, U., Rothe, T., 2013. Doing Well in Reforming the Labour Market? Recent Trends in Job Stability and Wages in Germany: IZA Discussion Paper No. 7580. Institute for the Study of Labor, Bonn.
Goedemé, T., 2013. How much confidence can we have in EU-SILC? Complex sample designs and the standard error of the Europe 2020 poverty indicators. Soc. Indic. Res. 110 (1), 89–110.
Gottschalk, P., Smeeding, T., 1997. Cross-national comparisons of earnings and income inequality. J. Econ. Lit. XXXV, 633–687.
Gottschalk, P., McLanahan, S.S., Sandefur, G.D., 1994. The dynamics and intergenerational transmission of poverty and welfare participation. In: Danziger, S.D., Sandefur, G.D., Weinberg, D.H. (Eds.), Confronting Poverty: Prescription for Change. Harvard University Press, Cambridge.
Gregg, P., Harkness, S., Machin, S., 1999. Child Development and Family Income. Joseph Rowntree Foundation, York.
Grogger, J., 2003. The effects of time limits, the EITC, and other policy changes on welfare use, work, and income among female-headed families. Rev. Econ. Stat. 85 (2), 394–408.
Grogger, J., 2004. Welfare transitions in the 1990s: the economy, welfare policy, and the EITC. J. Policy Anal. Manage. 23 (4), 671–695.

Guio, A.-C., 2009. What can be learned from deprivation indicators in Europe?: Eurostat Methodologies and Working Paper. Eurostat, Luxembourg.
Guio, A.-C., Maquet, E., 2007. Material deprivation and poor housing. Comparative EU Statistics on Income and Living Conditions: Issues and Challenge. Office for Official Publications of the European Communities, Eurostat, Luxembourg.
Hanushek, E., Woessmann, L., 2006. Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. Econ. J. 116 (510), C63–C76, 03.
Harding, A., Warren, N., Lloyd, R., 2006. Moving beyond traditional cash measures of economic well-being: including indirect benefits and indirect taxes: NATSEM Discussion Papers, Issue 61/2006. NATSEM, Canberra.
Heckman, J.J., 2010. Building bridges between structural and program evaluation approaches to evaluating policy. J. Econ. Lit. 48 (2), 356–398.
Heckman, J., Lochner, L., 2000. Rethinking education and training policy. In: Danziger, S., Waldfogel, J. (Eds.), Securing the Future. Russell Sage Foundation, New York.
Heckman, J.J., Urzúa, S., 2010. Comparing IV with structural models: what simple IV can and cannot identify. J. Econ. 156 (1), 27–37.
Heckman, J., Lalonde, R., Smith, J., 1999. The economics and econometrics of active labor market programs. In: Ashenfelter, O., Car, D. (Eds.), In: Handbook of Labour Economics, vol. 3. Elsevier/North Holland.
Hemerijck, A., Vandenbroucke, F., 2012. Social investment and the euro crisis: the necessity of a unifying social policy concept. Intereconomics 47 (4), 200–206.
Hills, J., 2002. Comprehensibility and balance: the case for putting indicators in baskets. Polit. Econ. 1, 95–98.
Holt, S., 2011. Ten Years of the EITC Movement: Making Work Pay Then and Now. Brookings Metropolitan Policy Program, Washington, DC.
Holzmann, R., Hinz, R., 2005. Old-Age Income Support in the 21st Century: The World Bank's Perspective on Pension Systems and Reform. The World Bank, Washington, DC.
Hotz, V.J., Scholz, J.K., 2003. The earned income tax credit. In: Moffit, R. (Ed.), Means-Tested Transfer Programs in the U.S. University of Chicago Press, Chicago, IL.
Hoynes, H., Page, M., Stevens, A., 2006. Poverty in America: trends and explanations. J. Econ. Perspect. 20, 47–68.
Imbens, G., 2010. Better LATE than nothing: some comments on Deaton (2009) and Heckman and Urzua (2009). J. Econ. Lit. 48 (2), 399–423.
Immervoll, H., 2012. Minimum-income benefits in OECD countries: policy design, effectiveness and challenges. In: Besharov, D., Couch, K. (Eds.), Measuring Poverty, Income Inequality, and Social Exclusion. Lessons from Europe. Oxford University Press, Oxford.
Immervoll, H., Jacobsen Kleven, H., Thustrup Kreiner, C., Saez, E., 2007. Welfare reform in European countries: a microsimulation analysis. Econ. J. 117 (516), 1–44.
Immervoll, H., Pearson, M., 2009. A good time for making work pay? Taking stock of in-work benefits and related measures across the OECD: OECD Social, Employment and Migration Working Papers No. 81. OECD, Paris.
Immervoll, H., Richardson, L., 2011. Redistribution policy and inequality reduction in OECD countries: what has changed in two decades? IZA Discussion Papers 6030. Institute for the Study of Labor (IZA).
Immervoll, H., Levy, H., Lietz, D., Mantovani, D., Sutherland, H., 2006. The sensitivity of poverty rates in the European Union to macro-level changes. Cambridge J. Econ. 30, 181–199.
Jäntti, M., Danziger, S., 2000. Income poverty in advanced countries. In: Atkinson, A.B., Bourguignon, F. (Eds.), Handbook of Income Distribution. Elsevier, Amsterdam.
Jäntti, M., Bratsberg, B., Roed, K., Raaum, O., Naylor, R., Österbacka, E., Björklund, A., Eriksson, T., 2006. American exceptionalism in a new light: a comparison of intergenerational earnings mobility in the Nordic countries, the United Kingdom and the United States: IZA Discussion Paper No. 1938. Institute for the Study of Labor, Bonn.
Jenkins, S.P., Brandolini, A., Micklewright, J., Nolan, B., 2013. The Great Recession and the Distribution of Household Income. Oxford University Press, Oxford.

Jones, A., O'Donnell, O., 1995. Equivalence scales and the costs of disability. J. Public Econ. 56 (2), 273–289.
Kangas, O., 1995. Attitudes to means-tested social benefits in Finland. Acta Sociol. 38, 299–310.
Kaplanoglou, G., Rapanos, V., Bardakas, I., 2013. Does fairness matter for the success of fiscal consolidation?: Working Paper, Available at SSRN: http://ssrn.com/abstract=2267831 or http://dx.doi.org/10.2139/ssrn.2267831.
Karoly, L., Kilburn, R., Cannon, J., 2005. Early Childhood Interventions. Rand Corporation, Santa Monica, CA.
Kenworthy, L., 2004. Egalitarian Capitalism? Jobs, Incomes and Inequality in Affluent Countries. Russell Sage Foundation, New York.
Kenworthy, L., 2008. Jobs with Equality. Oxford University Press, Oxford.
Kenworthy, L., 2011. Progress for the Poor. Oxford University Press, Oxford.
Kluve, J., 2010. The effectiveness of European active labor market programs. Labour Econ. 17 (6), 904–918.
Korpi, W., Palme, J., 1998. The paradox of redistribution and strategies of equality: welfare state institutions, inequality, and poverty in the western countries. Am. Sociol. Rev. 63 (5), 661–687.
Kraus, M., 2004. Social security strategies and redistributive effects in European social transfer systems. Rev. Income Wealth 50 (3), 431–457.
Kronauer, M., 1998. 'Social Exclusion' and 'Underclass'—new concepts for the analysis of poverty. In: Andreß, H.-J. (Ed.), Empirical Poverty Research in Comparative Perspective. Ashgate, Aldershot.
Lampman, R., 1971. Ends and Means of Reducing Income Poverty. Markham, Chicago.
Lane, P.R., Milesi-Ferretti, G.M., 2012. External adjustment and the global crisis. J. Int. Econ. 88 (2), 252–265, Elsevier.
Layte, R., Whelan, C.T., 2002. Cumulative disadvantage or individualization: a comparative analysis of poverty risk and incidence. Eur. Soc. 4 (2), 209–223.
Lefebvre, M., 2007. The redistributive effects of pension systems in Europe: a survey of evidence: LIS Working Paper Series. Luxembourg Income Study (LIS) No. 457.
Leigh, A., 2010. Who benefits from the earned income tax credit? Incidence among recipients, coworkers and firms. B.E. J. Econ. Anal. Policy, 10 (1), Berkeley Electronic Press.
Levine, R.A., Watts, H., Hollister, R., Williams, W., O'Connor, A., Widerquist, K., 2005. A retrospective on the negative income tax experiments: looking back at the most innovative field studies in social policy. In: Widerquist, K. (Ed.), The Ethics and Economics of the Basic Income Guarantee. Ashgate, Aldershot.
Liebman, J.B., 2002. Redistribution in the current U.S. social security system. In: Feldstein, M., Liebman, J.B. (Eds.), The Distributional Aspects of Social Security and Social Security Reform. University of Chicago Press, Chicago.
Lohmann, H., 2009. Welfare states, labour market institutions and the working poor: a comparative analysis of 20 European countries. Eur. Soc. Rev. 25 (4), 26.
Lupton, R., 2003. Neighbourhood Effects: Can We Measure Them and Does It Matter?: CASE Paper 73. Centre for the Analysis of Social Exclusion, London School of Economics, London.
Machin, S., 2008. Education and inequality. In: Salverda, W., Nolan, B., Smeeding, T. (Eds.), The Oxford Handbook of Economic Inequality. Oxford University Press, Oxford.
Mahler, V.A., Jesuit, D.K., September, 2006. Fiscal redistribution in the developed countries: new insights from the Luxembourg Income Study. Soc. Econ. Rev. 4, 483–511.
Maitre, B., Nolan, B., Whelan, C.T., 2012. Low Pay, In-Work Poverty and Economic Vulnerability: A Comparative Analysis Using EU-SILC. Manchester School 80 (1), 99–116.
Marchal, S., Marx, I., Van Mechelen, N., 2014. The great-wake up call? Social citizenship and minimum income provisions in times of crisis. J. Soc. Policy 43 (2), 247–267.
Marical, F., Mira d'Ercole, M., Vaalavuo, M., Verbist, G., 2006. Publicly-Provided Services and the Distribution of Resources. OECD Social, Employment and Migration Working Papers No. 45, OECD, Paris.
Marical, F., Mira d'Ercole, M., Vaalavuo, M., Verbist, G., 2008. Publicly-provided services and the distribution of households' economics resources. OECD Econ. Stud. 44 (1), 9–47.
Marlier, E., Atkinson, A.B., Cantillon, B., Nolan, B., 2007. The EU and Social Inclusion: Facing the Challenges. Policy Press, Bristol.

Marx, I., Nelson, K., 2013. Minimum Income Protection in Flux. Palgrave Macmillan, Basingstoke.
Marx, I., Nolan, B., 2013. In-work poverty. In: Cantillon, B., Vandenbroucke, F. (Eds.), Reconciling Work and Poverty Reduction: How Successful Are European Welfare States? Oxford University Press, Oxford.
Marx, I., Vanhille, J., Verbist, G., 2012a. Combating in-work poverty in continental Europe: an investigation using the Belgian case. J. Soc. Policy 41 (1), 19–41.
Marx, I., Vandenbroucke, P., Verbist, G., 2012b. Will rising employment levels bring lower poverty: regression based simulations of the Europe 2020 target. J. Eur. Soc. Policy 22 (5), 472–486.
Marx, I., Marchal, S., Nolan, B., 2013a. Mind the gap: net incomes of minimum wage workers in the EU and the US. In: Marx, I., Nelson, K. (Eds.), Minimum Income Protection in the Flux. Palgrave MacMillan, Basingstoke.
Marx, I., Salanauskaite, L., Verbist, G., 2013b. The paradox of redistribution revisited, and that it may rest in peace?: IZA Discussion Paper Series, vol. 7414. Institute for the Study of Labor, Bonn.
Matsaganis, M., Leventi, C., 2013. The distributional impact of the Greek crisis in 2010. Fisc. Stud. 34 (1), 83–108.
Matsaganis, M., et al., 2005. Child poverty and family transfers in Southern Europe: IZA Discussion Paper Series, No. 1509. Institute for the Study of Labor, Bonn.
Maurin, E., 2002. The impact of parental income on early school transitions. J. Public Econ. 85 (3), 301–332.
Mayer, S., 1997. What Money Can't Buy. Harvard University Press, Cambridge, MA.
Mayer, S., Lopoo, L., 2008. Government spending and intergenerational mobility. J. Public Econ. 92 (1–2), 139–158.
McCarty, N., Pontusson, J., 2009. The political economy of inequality and redistribution. In: Salverda, W., Nolan, B., Smeeding, T. (Eds.), The Oxford Handbook of Economic Inequality. Oxford University Press, Oxford, 2008.
McGarry, K., Davenport, A., 1998. Pensions and the distribution of wealth. In: Wise, D.A. (Ed.), Frontiers in the Economics of Aging. University of Chicago Press, Chicago.
McLanahan, S., Sandefur, G., 1994. Growing Up with a Single Parent. Harvard University Press, Cambridge.
Medgyesi, M., Temesváry Z., 2013. Conditional cash transfers in high-income OECD countries and their effects on human capital accumulation. GINI Discussion Paper #84, Amsterdam.
Meyer, D., Wallace, G., 2009. Poverty levels and trends in comparative perspective. Focus 26 (2), 7–13, Madison: Institute for Research on Poverty.
Mincer, J., 1958. Investment in human capital and personal income distribution. J. Polit. Econ. 66 (4), 281–302.
Moene, K., Wallerstein, M., 2001. Inequality, social insurance and redistribution. Am. Polit. Sci. Rev. 95 (4), 859–874.
Moene, K., Wallerstein, M., 2003. Earnings inequality and welfare spending: a disaggregated analysis. World Polit. 55 (4), 485–516.
Moffit, R., 1989. Estimating the value of an in-kind transfer: the case of food stamps. Econometrica 57 (2), 385–409.
Moller, S., Huber, E., Stephens, J., Bradley, D., Nielsen, F., 2003. Determinants of relative poverty in advanced capitalist democracies. Am. Sociol. Rev. 68 (1), 22–51.
Morel, N., Palier, B., Palme, J., 2011. Towards a Social Investment Welfare State? Ideas, Policies and Challenges. Policy Press, Bristol.
Myles, J., 2000. Poverty indices and poverty analysis. Rev. Income Wealth 46, 161–179.
Nelson, K., 2004. Mechanisms of poverty alleviation: anti-poverty effects of non-means tested and means-tested benefits in five welfare states. J. Eur. Soc. Policy 14 (1), 371–390.
Nelson, K., 2007. Universalism versus targeting: the vulnerability of social insurance and means-tested minimum income protection in 18 countries, 1990–2002. Int. Soc. Sec. Rev. 60, 33–58.
Nolan, B., 2013. What use is social investment? J. Eur. Soc. Policy 23 (5), 459–468.
Nolan, B., Marx, I., 2000. Low pay and household poverty. In: Gregory, M., Salverda, W., Bazen, S. (Eds.), Labour Market Inequalities: Problems and Policies of Low-Wage Employment in International Perspective. Oxford University Press, Oxford, pp. 100–119.
Nolan, B., Marx, I., 2009. Inequality, poverty and social exclusion. In: Salverda, W., Nolan, B., Smeeding, T. (Eds.), Oxford Handbook of Economic Inequality. Oxford University Press, Oxford.
Nolan, B., Russell, H., 2001. Non-Cash Benefits and Poverty in Ireland: Policy Research Series Paper No. 39. The Economic and Social Research Institute, Dublin.

Nolan, B., Whelan, C.T., 1996. Resources, Deprivation and Poverty. Clarendon Press, Oxford.
Nolan, B., Whelan, C.T., 2010. Using non-monetary deprivation indicators to analyze poverty and social exclusion: lessons from Europe? J. Policy Anal. Manage. 29 (2), 305–325.
Nolan, B., Whelan, C.T., 2011. Poverty and Deprivation in Europe. Oxford University Press, Oxford.
Nolan, B., Esping-Andersen, G., Whelan, C.T., Maitre, B., Wagner, S., 2011. The role of social institutions in intergenerational mobility. In: Erikson, R., Jantti, M., Smeeding, T. (Eds.), Persistence, Privilege, Policy and Parenting: The Comparative Study of Intergenerational Mobility. Russell Sage Foundation, New York.
OECD, 2001. When Money is Tight: Poverty Dynamics in OECD Countries. Chapter 3 of Economic Outlook, OECD, Paris.
OECD, 2005. Labour Market Programmes and Activation Strategies: Evaluating the Impacts: OECD Employment Outlook 2005. OECD, Paris (Chapter 4).
OECD, 2007a. Benefits and Wages. OECD, Paris.
OECD, 2007b. Activating the Unemployed: What Countries Do: OECD Employment Outlook 2007. OECD, Paris (Chapter 5).
OECD, 2008. Growing Unequal: Income Distribution and Poverty in OECD Countries. OECD, Paris.
OECD, 2009. Employment Outlook: Tackling the Jobs Crisis: Is Work the Best Antidote to Poverty? OECD, Paris, pp. 165–210 (Chapter 3).
OECD, 2011a. Divided We Stand. Why Inequality Is Rising. OECD, Paris.
OECD, 2011b. Pensions at a Glance 2011. Retirement Income Systems in OECD and G20 Countries. OECD, Paris.
OECD, 2013. Activating Jobseekers: Lessons from Seven OECD Countries: OECD Employment Outlook 2013. OECD, Paris.
Olivera, J., 2012. Preferences for Redistribution in Europe: Working Papers 2012/25. Geary Institute, University College Dublin.
Page, M.E., 2004. New evidence on the intergenerational correlation in welfare participation. In: Corak, M. (Ed.), Generational Income Mobility in North America and Europe. Cambridge University Press, pp. 226–244.
Paulus, A., Sutherland, H., Tsakloglou, P., 2010. The distributional impact of in-kind public benefits in European countries. J. Policy Anal. Manage. 29 (2), 243–266.
Pestieau, P., 2006. The Welfare State in the European Union. Oxford University Press, Oxford.
Piketty, T., 2003. The impact of divorce on school performance. Evidence from France, 1968–2002: CEPR Discussion Paper 4146. CEPR, London.
Raaum, O., Salvanes, K., Sørensen, E., 2003. The impact of a primary school reform on educational stratification: a Norwegian study of neighbour and school mate correlations. Swedish Econ. Policy Rev. 10 (2), 143–169.
Raaum, O., Bratsberg, B., Røed, K., Österbacka, E., Eriksson, T., Jäntti, M., Naylor, R., 2007. Marital sorting, household labor supply, and intergenerational earnings mobility across countries. B.E. J. Econ. Anal. Policy 7 (2) (Article 7).
Rothstein, J., 2010. Is the EITC as good as an NIT? Conditional cash transfers and tax incidence. Am. Econ. J. Econ. Policy 2 (1), 177–208.
Sanbonmatsu, L., Kling, J., Duncan, G., Brooks-Gunne, J., 2006. Neighborhoods and academic achievement: results from the moving to opportunity experiment. J. Hum. Res. 41 (4), 649–691.
Schütz, G., Ursprung, H., Woessmann, L., 2005. Education policy and equality of opportunity: IZA Discussion Paper No. 1906. Institute for the Study of Labor, Bonn.
Sen, A., 1976. Poverty: an ordinal approach to measurement. Econometrica 44, 219–231.
Sen, A., 1980. Equality of what. In: McMurrin, S.M. (Ed.), Tanner Lectures in Human Values I. Cambridge University Press, Cambridge.
Sen, A., 1993. Capability and well-being. In: Nussbaum, M., Sen, A. (Eds.), The Quality of Life. Oxford University Press, Oxford.
Sigle-Rushton, W., 2004. Intergenerational and life-course transmission of social exclusion in the 1970 British Cohort Study: Centre for Analysis of Social Exclusion Discussion Paper No. 78. London School of Economics and Political Science, London.
Skocpol, T., 1991. Targeting within universalism: politically viable policies to combat poverty in the United States. In: Jencks, C., Peterson, P.E. (Eds.), The Urban Underclass. The Brookings Institution, Washington, DC, pp. 411–436.

Smeeding, T., Sandstrom, S., 2005. Poverty and income maintenance in old age: a cross-national view of low income older women. Fem. Econ. 11 (2), 163–174.
Smeeding, T., Thompson, J., 2013. Inequality and Poverty in the United States: the Aftermath of the Great Recession, Finance and Economics Discussion Series 2013–51. Federal Reserve Board, Washington.
Smeeding, T., Williamson, J., 2001. Income maintenance in old age: what can be learned from cross-national comparisons: LIS Working Paper No. 263. Luxembourg Income Study, Luxembourg.
Smeeding, T.M., Saunders, P., Coder, J., Jenkins, S.P., Fritzell, J., Hagenaars, A.J.M., Hauser, R., Wolfson, M., 1993. Poverty, inequality, and family living standards impacts across seven nations: the effect of noncash subsidies for health, education, and housing. Rev. Income Wealth 39, 229–256.
Solon, G., Page, M., Duncan, G., 2000. Correlations between neighboring children in their subsequent educational attainment. Rev. Econ. Stat. 82 (3), 383–392.
Tóth, I., 2014. Revisiting grand narratives of growing inequalities: lessons from 30 country studies. In: Nolan, B., Salverda, W., Checchi, D., Marx, I., McKnight, A., Tóth, I.G., van de Werfhorst, H. (Eds.), Changing Inequalities and Societal Impacts in Rich Countries: Thirty Countries' Experiences. Oxford University Press, Oxford.
Townsend, P., 1979. Poverty in the United Kingdom. Penguin, Harmondsworth.
Tsakloglou, P., Papadopoulos, F., 2002. Aggregate level and determining factors of social exclusion in twelve European countries. J. Eur. Soc. Policy 12 (3), 209–223.
Tsui, K., 2002. Multidimensional poverty indices. Soc. Choice Welfare 19, 69–93.
UNICEF, 2007. Child poverty in perspective: an overview of child well-being in rich countries: Innocenti Report Card 7. UNICEF, Florence, Italy.
US Census Bureau, 2003. Dynamics of Economic Well-Being: Poverty 1996–1999. US Census Bureau, Washington, DC.
Valletta, R., 2006. The ins and outs of poverty in advanced economies: government policy and poverty dynamics in Canada, Germany, Great Britain, and the United States. Rev. Income Wealth 52, 261–284.
van Kersbergen, K., Hemerijck, A., 2012. Two decades of change in Europe: the emergence of the social investment state. J. Soc. Policy 41, 475–492.
Van Lancker, W., Ghysels, J., 2012. Who benefits? The social distribution of subsidized childcare in Sweden and Flanders. Acta Sociol. 55, 125–142.
Van Mechelen, N., Bradshaw, J., 2013. Child poverty as a government priority: child benefit packages for working families, 1992–2009. In: Marx, I., Nelson, K. (Eds.), Minimum Income Protection in Flux. Palgrave Macmillan, Basingstoke.
Van Mechelen, N., Marchal, S., Goedemé, T., Marx, I., Cantillon, B., 2011. The CSB Minimum Income Protection Indicators dataset (CSB-MIPI). CSB Working Paper, Herman Deleeck Centre for Social Policy, Antwerp.
Van Mechelen, N., Marchal, S., 2013. Struggle for life: social assistance benefits, 1992–2009. In: Marx, I., Nelson, K. (Eds.), Minimum Income Protection in Flux. Palgrave Macmillan, Basingstoke.
Van Oorschot, W., 2002. Targeting welfare: on the functions and dysfunctions of means-testing in social policy. In: Townsend, P., Gordo, D. (Eds.), World Poverty: New Policies to Defeat an Old Enemy. The Policy Press, Bristol.
Van Vliet, O., Been, J., Caminada, K., Goudswaard, K., 2012. Pension reform and income inequality among older people in 15 European countries. Int. J. Soc. Welf. 21, S8–S29.
Vandenbroucke, F., Vleminckx, K., 2011. Disappointing poverty trends: is the social investment state to blame? J. Eur. Soc. Policy 21, 432–449.
Vandenbroucke, F., Hemerijck, A., Palier, B., 2011. The EU Needs a Social Investment Pact: OSE Paper Series, Opinion Paper No. 5. Observatoire Social Europeene, Brussels.
Vandenbroucke, F., Cantillon, B., Van Mechelen, N., Goedemé, T., Van Lancker, A., 2013. The EU and minimum income protection: clarifying the policy conundrum. In: Marx, I., Nelson, K. (Eds.), Minimum Income Protection in Flux. Palgrave MacMillan, Basingstoke.
Verbist, G., 2004. Redistributive effect and progressivity of income taxes: an international comparison across the EU using EUROMOD: EUROMOD Working Paper EM5/04. Microsimulation Unit, Cambridge University, Cambridge.

Verbist, G., Figari, F., 2014. The redistributive effect and progressivity of taxes revisited: an international comparison across the European Union. FinanzArchiv 70, 1–25.
Vignoli, D., De Santis, G., 2010. Individual and contextual correlates of economic difficulties in old age in Europe. Popul. Res. Policy Rev. 29, 481–501.
Waldfogel, J., 2006. What Children Need. Harvard University Press, Cambridge, MA.
Whelan, C.T., Maitre, B., 2010. Welfare regime and social class variation in poverty and economic vulnerability in Europe: an analysis of EU-SILC. J. Eur. Soc. Policy 20 (4), 316–332.
Whelan, C.T., Layte, R., Maître, B., Nolan, B., 2001. Income, deprivation and economic strain: an analysis of the European community household panel. Eur. Soc. Rev. 17 (4), 357–372.
Whelan, C.T., Layte, R., Maître, B., 2003. Persistent income poverty and deprivation in the European Union. J. Soc. Policy 32 (1), 1–18.
Whelan, C.T., Nolan, B., Maître, B., 2013. Analysing intergenerational influences on income poverty and economic vulnerability with EU-SILC. Eur. Soc. 15 (1), 82–105.
Whiteford, P., 2008. How much redistribution do governments achieve? The role of cash transfers and household taxes. In: OECD (Ed.), Growing Unequal? OECD, Paris.
Whiteford, P., Adema, W., 2007. What works best in reducing child poverty: a benefit or work strategy?: OECD Social Employment and Migration Working Papers, No. 51. OECD, Paris.
Wilson, W.J., 1987. The Truly Disadvantaged: The Inner City, The Underclass and Public Policy. University of Chicago Press, Chicago.
Wolff, E., Zacharias, A., 2007. The distributional consequences of government spending and taxation in the U.S., 1989 and 2000. Rev. Income Wealth 53 (4), 692–715.
Zaidi, A., Burchardt, T., 2005. Comparing incomes when needs differ: equivalization for the extra costs of disability in the U.K. Rev. Income Wealth 51 (1), 89–114.
Zaidi, A., Makovec, M., Fuchs, M., Lipszyc, B., Lelkes, O., Grech, A., Marin, B., de Vos, K., 2006a. Poverty of elderly people in EU25: Report submitted to the European Commission. European Centre for Social Welfare Policy and Research, Vienna.
Zaidi, A., Marin, B., Fuchs, M., 2006b. Pension policy in EU25 and its possible impact on elderly poverty and appendices: Report Submitted to the European Commission. European Centre for Social Welfare Policy and Research, Vienna.
Zheng, B., 2000. Poverty orderings. J. Econ. Surv. 14, 427–466.

CHAPTER 24

Microsimulation and Policy Analysis

Francesco Figari*,†, Alari Paulus†, Holly Sutherland†
*University of Insubria, Varese, Italy
†Institute for Social and Economic Research (ISER), University of Essex, Colchester, UK

Contents

Handbook of Income Distribution, Volume 2B
ISSN 1574-0056, http://dx.doi.org/10.1016/B978-0-444-59429-7.00025-X

Abstract

We provide an overview of microsimulation approaches for assessing the effects of policy on income distribution. We focus on the role of tax-benefit policies and review the concept of microsimulation and how it contributes to the analysis of income distribution in general and policy evaluation in particular. We consider the main challenges and limitations of this approach and discuss directions for future developments.

Keywords

Microsimulation models, Income distribution, Tax-benefit policies

JEL Classification Codes

C81, D31, H30, I30

24.1. INTRODUCTION AND OVERVIEW

24.1.1 What is Microsimulation?

Microsimulation methods are increasingly used to evaluate the effects of policies on income distribution. Microsimulation refers to a wide variety of modeling techniques that operate at the level of individual units (such as persons, firms, or vehicles), with rules applied to simulate changes in state or behavior. These rules may be deterministic or stochastic, with the result being an estimate of the outcomes of applying these rules, possibly over many steps involving many interactions. These estimates are also at the micro level, allowing analysis of the distribution of the outcomes and changes to them, as well as the calculation of any relevant aggregate.[1]

In the social sciences, Guy Orcutt and his colleagues pioneered microsimulation models in the 1950s as a new approach to analyzing the impact of social and economic policies, which accounted for the characteristics and behavior of the microlevel units

[1] Adapted from the definition provided by the International Microsimulation Association (http://microsimulation.org/). Broadly speaking, microsimulation modeling could also cover agent-based simulation (ABS), though they have remained very distinct fields in the literature with microsimulation methods drawing heavily on micro-data (Spielauer, 2011).

under investigation (Orcutt, 1957; Orcutt et al., 1961). Microsimulation is commonly applied to many areas relevant to public policy, such as transportation, location planning for public services and commercial developments, and demand for health care and long-term care.[2] The microsimulation approaches considered here are those that primarily address questions related to the impact of tax-benefit policies on income distribution. Models simulating the effects of social and fiscal policies on household income were first developed in the 1980s when the essential inputs—micro-data from household surveys and accessible computing power—began to be made available.

These early tax-benefit microsimulation models were arithmetic, recalculating the components of household disposable income (usually cash benefits, direct taxes, and social contributions) for each household in a representative microdataset under different sets of policy rules. They could answer "what if" questions about the effects of specific policy reforms on each household's income and hence on the overall income distribution and the aggregate public budget. Some early studies include Atkinson et al. (1983) and Betson et al. (1982). These models could also readily be used to calculate indicators of work incentives on the intensive margin (Atkinson and Sutherland, 1989; Bourguignon et al., 1993). Since then, this "static" modeling approach has not only proliferated, but it has also been refined in a number of directions, influenced by developments in data availability, methodology, speed, capacity of accessible computing power, and the demands made by policymaking and policy analysis.

Microsimulation models are often categorized as "static," "dynamic," or "behavioral" (see Harding, 1996a). The first type applies purely deterministic policy rules to microdata in combination with data adjustments such as reweighting. The characteristics of the micro units stay constant. Dynamic models, on the other hand, "age" the micro units through time, changing their characteristics in response to natural processes and the probabilities of relevant events and transitions (Li and O'Donoghoue, 2013). Behavioral models use microeconometric models of individual preferences to estimate the effects of policy changes on behavior, often in terms of labor supply. In practice the distinction between modeling approaches is no longer necessarily useful because modern microsimulation analysis often combines elements of each type, according to the question being addressed. For example, labor supply models require the calculation of budget sets (household income under alternative labor supply scenarios) for individuals, and these are usually generated by static tax-benefit models. Behavioral reactions, as well as static calculations, are relevant in dynamic microsimulations. In seeking to simulate the effects of policy changes in a variety of economic environments, so-called static models may borrow elements from dynamic model methodology, and in seeking to simplify the dynamic modeling process, the reverse can also be true (Caldwell, 1990). In practice,

[2] There are extensive literatures covering each area. See, for example, Dowling et al. (2004), Waddell et al. (2003), and Gupta and Harding (2007).

dynamic models mainly address questions about the effects of policies that take time to evolve, such as pensions (e.g., Borella and Coda Moscarola, 2010; Dekkers et al., 2010; Flood, 2007) and long-term care reform (e.g., Hancock, 2000; Hancock et al., 2013), often focusing on the cost, winners, and losers, as well as questions about intrapersonal redistribution over the lifecycle (Harding, 1993).

Without tax-benefit microsimulation modeling, and before it was widespread, analysis of the effects of taxes and benefits on household income, as well as calculation of work incentive indicators, was limited to "model family" calculations for stylized households, sometimes referred to as "tax-benefit models." These calculations are carried out, for example, by the Organization for Economic Cooperation and Development (OECD) for the purpose of making cross-country comparisons (OECD, 2007), but, depite being useful for understanding the net effects of policies in particular standardized cases, such models cannot give full information about impacts on income distribution.

This chapter provides an overview of microsimulation approaches for exploring the effects of policy on income distribution, and it highlights some particular state-of-the-art or innovative studies that have been carried out in this area. The main emphasis is on static modeling methods, though we also consider extensions accounting for behavioral reactions (Section 24.3.3) and highlight the main modeling features of dynamic modeling (Section 24.5.2), referring to the existing reviews. We have not attempted to create a comprehensive review of the models themselves. Their proliferation would make such a task not only daunting, but quickly out-of-date. There are already a number of reviews and collections describing the models and the analyses using them, a selection of which we summarize below.

24.1.2 Microsimulation in the Economic Literature

There are several distinct motivations for using a microsimulation model to simulate the impact of a given policy on income distribution. Microsimulation can be used to quantify the role of *existing* policies on income inequality or poverty in a given context. More importantly, it is a tool to aid the design of new policies with particular objectives and to evaluate actual or proposed *reforms* in dimensions that were not taken into account in the original design. Moreover, it can also be used to show how alternative approaches could result in better outcomes in some respect. From a practical policy perspective, one of the main uses of microsimulation modeling for the design of policy is to assess the approximate budgetary cost of a new policy given its objectives, such as the desire to reduce the poverty gap or to increase work incentives for particular groups. Such analysis rarely sees the light of day except in its final form as a costed reform proposal.

Evidence from microsimulation modeling is also used to inform academic economic debates about the impact of policy reforms and the optimal design of policy (Blundell, 2012). In general terms, a microsimulation approach allows the researcher to conduct a controlled experiment by changing the parameters of interest while holding everything else constant and avoiding endogeneity problems in identifying the direct effects of the policy under analysis (Bourguignon and Spadaro, 2006). The use of tax-benefit microsimulation models to calculate counterfactual states and scenarios underpins much microeconomic analysis of the causal impact of fiscal policy reforms. A prime example is the use of the Institute of Fiscal Studies tax-benefit microsimulation model, TAXBEN for the UK, to provide empirical evidence for the arguments about tax design put forward in the authoritative Mirrlees Review (Mirrlees et al., 2010). Moreover, the counterfactuals shed light on the potential ingredients of optimal tax analysis, which cannot be derived in a quasiexperimental setting. This is demonstrated by the developments of the computational optimal income taxation theory, applied by Aaberge and Colombino (2013) to Norway and by Blundell and Shepard (2012) to the UK.

Microsimulation modeling is increasingly recognized as part of the policy evaluation literature, in which it is one of the key ingredients of a careful, evidence-based evaluation of the design of tax-benefit reforms. Although, in general, this literature has been more focused on ex-post analysis, Keane (2010) and Blundell (2012), among others, have underlined the need to consider both ex-ante and ex-post approaches to study the effects of policy changes. In this context, tax-benefit microsimulation models can offer insights in two ways. First, they are unique tools for conducting ex-ante analysis through the simulation of counterfactual scenarios reflecting alternative policy regimes. Such counterfactuals are needed both for the "morning-after" evaluation of tax-benefit reforms and for more complex structural models that reveal individual behavioral changes based on simulated budget constraints and an estimated model of individual and family choices (see Section 24.3.3). Second, by developing a counterfactual scenario, tax-benefit microsimulation models enable the researcher to disentangle ex-post what would have happened without a given policy. Although ex-post analysis is typically conducted by means of quasiexperimental approaches, based on difference-in-difference, matching, and selection estimators, the cross-fertilization between ex-ante and ex-post approaches has contributed to the increasing credibility of analysis based on detailed microsimulation models, making them a core part of the causal policy evaluation literature. A prime example is the quasiexperimental analysis used to validate structural models of labor supply that use microsimulation models to derive the budget sets faced by individuals (see, among others, Blundell, 2006).

Furthermore, microsimulation features in the strain of literature that involves micro–macro linkage, aiming to measure the effects of macroeconomic changes (including macroeconomic policy) on income distribution. More specifically, the linkage of microsimulation models to macroeconomic models allows one to consider the interactions of macroeconomic policies or shocks with the tax-benefit systems (see Section 24.3.4). Ignoring the tax-benefit policy effects on income distribution can be justifiable in some circumstances, for example, when analyzing their impact in developing countries, because they may be very limited in size, and reform to social expenditures or macroeconomic shocks could be much more relevant for redistribution, but it is more problematic in the context of mature welfare states (Bourguignon and Bussolo, 2013).

The literature on microsimulation has expanded enormously in the last 20 years, along with the spread and development of this methodology. An attempt to cover all relevant publications would be a daunting task, and, therefore, we aim to provide some of the most important methodological references with relevant illustrations in the rest of this chapter. For further and broader material, we refer the reader to a number of reviews and workshop and conference volumes that provide good surveys, both of model applications and of models themselves, reflecting how state-of-the-art modeling has evolved since the beginning of the 1990s: Harding (1996b), Gupta and Kapur (2000), Mitton et al. (2000), Gupta and Harding (2007), Harding and Gupta (2007), Lelkes and Sutherland (2009), Zaidi et al. (2009), Dekkers et al. (2014), and O'Donoghue (2014).[3] For surveys of the models themselves, see Merz (1991), Sutherland (1995), Klevmarken (1997), Gupta and Kapur (2000), O'Donoghue (2001), Zaidi and Rake (2001), Gupta and Harding (2007), Urzúa (2012), and Li and O'Donoghoue (2013). In addition, several books focus on specific models, providing excellent examples of opening the "black box" often associated with complex economic models. For example, Harding (1993) describes the details of her dynamic cohort microsimulation model used to evaluate lifetime income distribution and redistribution for Australia; Redmond et al. (1998) provide an extensive discussion of the inner workings of POLIMOD, a static tax-benefit model for the UK; and Bargain (2007) offers a collection of applications using EUROMOD, the EU tax-benefit model. Furthermore, the microsimulation community established the International Microsimulation Association (IMA) in 2005, and since 2007, it has been possible to follow the latest developments in the field through the *International Journal of Microsimulation*, a refereed online journal published under the auspices of the IMA.[4]

24.1.3 Summary of Chapter

The remainder of this chapter is structured as follows. Before getting into the ways in which microsimulation can be used to understand the effects of policy changes, Section 24.2 describes how it can be used to improve the information generally available for the analysis of income distribution and redistribution. Simulated estimates of tax liability and benefit entitlement can be used alongside the values recorded in survey and administrative microdatasets to understand and improve on the deficiencies in the latter (e.g., to impute gross income from net if the former is not available or measured satisfactorily in the source data). Furthermore, indicators that cannot be collected in surveys or through administrative processes but are of value in understanding the

[3] For older conference volumes and reviews, see literature references given in these collections.
[4] See http://www.microsimulation.org/ijm/.

relationships between policy and income distribution, such as indicators of work incentives, can be calculated using microsimulation models.

Throughout the chapter we provide some empirical illustrations drawing mainly on analysis using the EU-wide tax-benefit model EUROMOD (Sutherland and Figari, 2013). Covering 27 countries and made generally accessible, this has become one of the most widely used models. We have chosen to highlight EUROMOD at least partly because it is generally available to use, and readers can reproduce, update, and extend the chapter's examples of analysis with relative ease. More information about EUROMOD is provided in Box 24.1.

Box 24.1 EUROMOD—A tax-benefit microsimulation model

EUROMOD is the tax-benefit microsimulation model of the European Union. It simulates individual and household tax liabilities and cash benefit entitlements according to the policy rules in place, and reforms to them, in each member state. It has two main distinguishing features. First, it covers many countries within the same framework, enabling a wide range of applications and comparability of results. Generally, EUROMOD is much more flexible than national microsimulation models in order to ensure consistency of results and transferability of tax-benefit system components across countries. Second, it is intended to be openly accessible: use is not restricted to the owners of the model. The calculations carried out by EUROMOD for any one country are in other respects quite typical of all tax-benefit microsimulation models, at least for developed countries. The description below is therefore generally applicable.

EUROMOD combines information on policy rules with detailed and nationally representative microdata on individual and household circumstances drawn from household income surveys and other data sources. The rules for each policy instrument are applied arithmetically to the characteristics of each individual, resulting in the amount of tax liability or benefit entitlement. For example, in the case of the simplest universal child benefit, the number of children within the eligible age range in the family is counted and the benefit amount per child is multiplied by this number to give the family's entitlement. Further issues complicate the calculation: "child" and "family" need to be defined, and the interaction of the child benefit amount with the rest of the tax-benefit system needs to be accounted for. This illustrative calculation is taken further in Appendix A by considering the effects of a change in policy.

The results of the calculations for each household are stored at the micro level and can be analyzed with any statistical software. At their simplest they may be weighted to population level, and the weighted change in income can be added up to provide an estimate of the budgetary effect of the policy change, or it can be analyzed in relation to any characteristics provided in the data: for example, to show the proportion of households gaining and losing by income quantile, region, or household type. The micro-outputs from alternative policy or labor market scenarios can also be used as the basis for calculating indicators of work incentives or for modeling changes in labor supply or other behavior.

Continued

EUROMOD aims to simulate as many of the tax and benefit components of household disposable income as possible, and generally, the following instruments are simulated: income taxes, social insurance contributions, family benefits, housing benefits, social assistance, and other income-related benefits. Instruments that are not simulated are taken directly from the data. These include most contributory benefits and pensions (due to the lack of information on previous employment and contribution history) and disability benefits (because of the need to know the nature and severity of the disability, which is also not present in the data).

EUROMOD input data for most countries are derived from the European Union Statistics on Income and Living Conditions (EU-SILC). In common with most sources of microdata used as input into microsimulation models, the EU-SILC was not designed for this purpose (Figari et al., 2007). A significant amount of preparation of the data, including imputing necessary information that is missing, needs to be done. For example, if gross income values are not directly recorded during the data collection operations and are imputed in an unsatisfactory way, a net-to-gross procedure is applied to the net income variables in order to derive the gross values used in the policy simulation.

EUROMOD includes some simple adjustments for the non-take-up of some benefits and evasion of taxes in some countries. In common with other adjustments and assumptions (e.g., the updating of nonsimulated incomes to a more recent point in time than the data income reference point) these can be changed or "switched off" by the user, depending on the analysis being done.

Baseline systems in EUROMOD have been validated and tested at the micro level (i.e., case-by-case validation) and the macro level. For each system simulated in EUROMOD, Country Reports are available on the EUROMOD web pages with background information on the tax-benefit system(s), a detailed description of all tax-benefit components simulated, a general overview of the input data, and an extended summary of the validation process.

For more information about EUROMOD and its applications, see the official website (https://www.iser.essex.ac.uk/euromod) and Sutherland and Figari (2013).

The primary motivation for building a tax-benefit microsimulation model is to be able to analyze the effects of policy changes on income distribution. Section 24.3 starts with a description of the basic process and explains the need to carry out microlevel calculations in order to capture the effects due to the complexity of tax-benefit systems. However, in any microsimulation analysis, the modeler must choose which dimensions to focus on and which to hold constant. Most studies do not set out their specific choices in formal terms. Section 24.3.2 provides a formal framework applicable to most tax-benefit microsimulation analyses. The following four subsections focus on some of the major and commonly applied extensions to the basic approach. Section 24.3.3 discusses how individual behavioral responses to policy changes are estimated and focuses on labor supply responses. This is followed in Section 24.3.4 by a review of the ways changes in income distribution can be linked to macroeconomic processes. Section 24.3.5 covers the use of microsimulation, in conjunction with macrolevel statistics or forecasts, to provide estimates of income distribution for periods beyond those covered by the latest microdata. These projections might be for the current situation (nowcasting) or sometime in the future (forecasting). Finally, Section 24.3.6 focuses on the ways in which microsimulation can be used to inform cross-country comparisons of the effects of policies.

Of course, there are many remaining challenges to providing estimates of the effects of policy and policy changes that can be used with confidence within policy analysis, and Section 24.4 considers three major ones. First, Section 24.4.1 considers the issues around reconciling the simulated income distribution and that measured using the original microdata (from surveys particularly, but also administrative sources). A major difference between the two distributions can undermine confidence in microsimulation results but has a number of interrelated causes, some of which can point to problems in survey data (e.g., income underreporting), and can be mitigated using information from simulations, and others that cannot (e.g., small and unrepresentative samples of high-income earners). Simulations can overestimate income if the non-take-up of benefits is not accounted for and also distorted if there is tax evasion. These issues, and how they may be accounted for in microsimulation models, are discussed in Section 24.4.2. Finally, it is important that the reliability of microsimulation estimates is possible to ascertain. This applies both in terms of how well point estimates match up to information from other sources (validation) and the need for statistical reliability indicators that can be applied to microsimulation estimates. Section 24.4.3 considers these issues.

Although the main focus of this chapter is the contribution to policy analysis of (direct) tax and (cash) benefit microsimulation of household incomes at the national level at a given point in time, Section 24.5 considers a somewhat broader scope, in some dimensions. Section 24.5.1 discusses a broadening of the outcome income measure to include the effect of noncash benefits and, particularly, indirect taxes. Section 24.5.2 reviews the main features of dynamic microsimulation models used in analyzing the long-term redistributive effects of policies and the incidence of tax-benefit systems over the lifetime rather than cross-sectionally at a point in time. Section 24.5.3 discusses the use of microsimulation to explore the effects of policies at a lower level than that of the nation (e.g., Spanish regions or US states) and at a higher level (e.g., the European Union or world regions such as southern Africa).

The final section concludes by first summarizing our view of the achievements of microsimulation for policy analysis to date and then by exploring the outlook for the future along two dimensions: the need for data improvements and methodological developments and the need to consider ways to organize development, maintenance, and access to microsimulation models for policy analysis purposes.

24.2. WHAT DOES MICROSIMULATION ADD TO ANALYSIS OF INCOME DISTRIBUTION AND REDISTRIBUTION?

24.2.1 Enriching Existing Microdata

Although the most obvious application of the microsimulation method is assessing the effects of tax-benefit policy *changes* on income distribution, it can be also useful for analyzing the existing income distribution and redistribution. Compared to research on income distribution directly utilizing only survey or administrative data, fiscal microsimulation can complement and improve such analysis by (i) adding further information, (ii) checking the consistency of the collected data, and (iii) allowing for greater flexibility with respect to the unit of analysis.

24.2.1.1 Adding Information

Simulations allow the generation of data that may be too difficult or expensive to collect directly or accurately from individuals. A common use of microsimulation in the processing of income survey data is deriving gross incomes from the net values that are collected, or vice versa. Compared to other methods such as statistical imputation, microsimulation accounts for the full details of the tax-benefit rules that are applicable for a given individual or household. Hence, it provides more accurate results, but may also require more effort to develop and keep up-to-date. Specific microsimulation routines are often built for this purpose. Among others, see Betti et al. (2011) on the Siena Microsimulation Model, which is used for conversions between net and gross income variables for several countries in the European Union Statistics on Income and Living Conditions (EU-SILC) survey, and Jenkins (2011) on the derivations of net income variables for the British Household Panel Survey (BHPS).

Such gross-to-net conversion routines naturally follow the logic of full-scale tax-benefit models, though they may still have notable differences. For example, tax-benefit models typically deal with the final tax liability (i.e., aiming to account for all tax concessions and considering the total taxable income), but taxes withheld on specific income sources are often more relevant for gross/net adjustments in a survey. For net-to-gross conversions, there are two microsimulation-related approaches. One is to apply inverted statutory tax rules and the other is to use gross-to-net routines in an iterative procedure to search for the corresponding gross value for a given net income, as suggested in Immervoll and O'Donoghue (2001). The first approach can be more straightforward if tax rules are relatively simple and analytical inversion is feasible, while the second approach allows the use of already existing tax-benefit models. The latter approach has also been used in the Siena model and related applications, as discussed by Rodrigues (2007).

If a tax-benefit model is applied to income data that contain imputed gross values, it is important to ensure that the net-to-gross conversion is consistent with the tax-benefit model calculations because otherwise simulated net incomes will not match the observed values. This source of bias is easy to overlook, and consistency is often difficult to establish because the documentation of net-to-gross derivations carried out by survey data providers often lacks sufficient details for tax-benefit modeling purposes.

Microsimulation methods can be also used to obtain more detailed tax information compared to what is usually available in the surveys (if any at all). For example, the Current Population Survey (CPS), one of the main household surveys in the US, provides such information through the Annual Social and Economic Supplement (ASEC), which includes simulated direct taxes and imputed employer's contributions for health insurance (Cleveland, 2005). Alternatively, surveys could be combined with detailed tax information from administrative records, though in practice this is still underdeveloped due to limitations on access to administrative records. Furthermore, microsimulation models can extend the scope of income information by simulating employer social insurance contributions and indirect taxes, which are usually not captured in income (and expenditure) surveys, even though their economic incidence is typically considered to be borne by individuals (Fullerton and Metcalf, 2002) and hence relevant for welfare analysis.

Although benefit information tends to be more detailed in income surveys, there are applications in which microsimulation methods can still provide further insights. Specifically, microsimulation allows the assessment of the intended effect of transfers (by calculating benefit eligibility) and contrasts it with reported outcome (i.e., observed benefit receipt), which is influenced by individual compliance behavior (see more in Section 24.4.2) and the effectiveness of benefit administrations, among other factors.

Of course, it is possible to carry out analysis of the redistributive effect of taxes and benefits only using survey information directly. For example, Mahler and Jesuit (2006), Immervoll and Richardson (2011) and Wang et al. (2012) use household survey data from the Luxembourg Income Study (LIS) to analyze redistributive effects in the OECD countries, and Fuest et al. (2010) and Atta-Darkua and Barnard (2010) use the SILC data for EU countries. However, microsimulation methods can often add to the scope and detail of the analysis. For example, Immervoll et al. (2006a), Paulus et al. (2009), Jara and Tumino (2013) use tax and benefit data simulated with EUROMOD for EU countries, and Kim and Lambert (2009) analyze redistribution in the US on the basis of the CPS/ASEC. Wagstaff et al. (1999) specifically analyze the progressivity of personal income taxes in the OECD countries also using the LIS data, and Verbist and Figari (2014) carry out similar analysis for EU countries, relying on EUROMOD simulations, allowing them to extend the analysis with social insurance contributions as well. Piketty and Saez (2007) use the TAXSIM[5] model to compute US federal individual income taxes and analyze their progressivity. Furthermore, Verbist (2007) employs EUROMOD to consider the distribution and redistributive effects of replacement incomes, taking into account interactions with taxes and social contributions, and Hungerford (2010) uses simulations to examine certain federal tax provisions and transfer programs in the US. Decoster and Van Camp (2001), O'Donoghue et al. (2004), and Decoster et al. (2010) are examples of studies simulating and analyzing the effects of indirect taxes across the distribution of income.

Microsimulation can additionally help to detect inconsistencies and potential measurement errors in the existing data. An obvious example is cross-checking whether gross and net income values (if both are reported) correspond to each other. As benefit income tends to be underreported in survey data (Lynn et al., 2012; Meyer et al., 2009), use of simulated benefits has the potential to improve the accuracy of income information (see more in Section 24.4.1). However, the quality of input data is also critical for the simulated results themselves, and there could be other reasons for discrepancies between observed and simulated income apart from underreporting (see Figari et al., 2012a).

24.2.1.2 The Unit of Analysis

Microsimulation can also offer some flexibility in the choice of unit of analysis. In any analysis of distribution, the unit of measurement is an important issue. Income is often measured at the household level, aggregating all sources across all individuals. Income surveys may not facilitate analysis at a lower level (e.g., aggregating within the narrow family or the fiscal unit) because some or all income variables are provided only at the household level. This is the case, for example, for the microdata provided by Eurostat from the European Statistics on Income and Living Conditions (EU-SILC). However, considering the effect of policy on the incomes of subunits within households may be relevant in a number of ways. The assumption of complete within-household sharing of resources deserves to be questioned, and its implications made clear. For example, assessments of poverty risk among pension recipients might look quite different if

[5] TAXSIM is the NBER microsimulation model that calculates US federal and state income taxes (http://www.nber.org/taxsim/). It covers the federal tax system from 1960 and the state systems from 1977 up to the current year. Model calculations are done in the TAXSIM server on the basis of survey data provided by the users in the required format containing different sources of income, deductions, and personal characteristics used to calculate tax liabilities. The program, written in FORTRAN, reads the input data sent by the user through a web application, calculates tax liabilities, and loads the results on the user's computer. Recent applications are based on the March Current Population Survey, the Survey of Consumer Finance, the Consumer Expenditure Survey, and the Panel Study of Income Dynamics, and a library of scripts used to derive the input data from different sources is made available by previous users. See Feenberg and Coutts (1993) for more information.

researchers did not assume that they shared this income with coresident younger generations, and vice versa. Furthermore, it may be particularly relevant to consider the effects of policy in terms of the particular unit of assessment, rather than the household as a whole. Minimum income schemes use a variety of units over which to assess income and eligibility, and these are often narrower than the survey household. A flexible microsimulation model is able to operate using a range of units of analysis, as well as units of assessment and aggregation, because they are able to assign income components, or shares of them, to the relevant recipient units within the household. Examples of microsimulation studies that consider units of analysis apart from the household are Decoster and Van Camp (2000) in relation to tax incidence at the household or fiscal unit level, Figari et al. (2011a) who analyze income within couples, and Bennett and Sutherland (2011) who consider the implications of means-testing at the family-unit level for receipt of benefit income by individuals.

24.2.2 Microsimulation-Based Indicators

The microsimulation method is also used to construct various indicators to measure the extent to which household disposable income reacts to changes in gross earnings or individual or household characteristics through interactions with the tax-benefit system. The two main groups of such indicators reflect individual work incentives and automatic adjustment mechanisms built into fiscal systems. This subsection gives an overview of these indicators and provides some examples, and a more formal presentation can be found in Section 24.3.2.

24.2.2.1 Indicators of Work Incentives

Marginal effective tax rates and participation tax rates are indicators of work incentives for the intensive (i.e., work effort) and the extensive labor supply margin (i.e., decision to work), respectively. Marginal effective tax rates (METR) reflect the financial incentive for a working person to increase his work contribution marginally either through longer hours or higher productivity (increasing the hourly wage rate). They show the proportion of additional earnings that is taxed away, taking into account not only the personal income tax but also social contributions as well as interactions with benefits, including withdrawal of means-tested benefits as private income increases. As such, METRs indicate more accurately the actual tax burden on additional income compared to statutory marginal income tax rates. Given that taxes and benefits form a complex nonlinear system, it is usually not feasible to obtain METRs in the form of analytical derivatives of the overall tax-benefit function. Instead, METRs are estimated empirically by incrementing gross earnings of an employed person by a small margin (e.g., 1–5%) and recalculating disposable income, as discussed by Immervoll (2004), Adam et al. (2006b), and Jara and Tumino (2013). Figure 24.1 provides an example from the last of these showing the extent to which average METRs and their distributions vary across the European

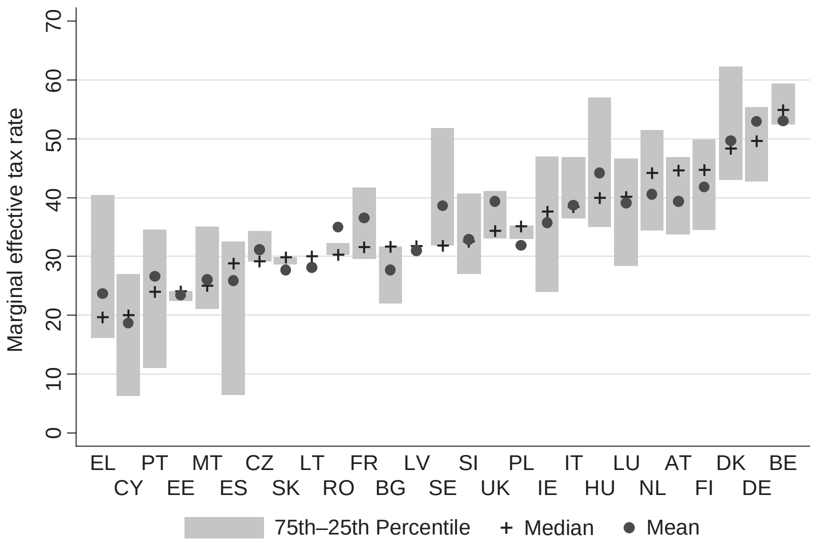


Figure 24.1 Marginal effective tax rates across the EU, 2007 (%). *Notes: Countries are ranked by median METR. Source: Jara and Tumino (2013), using EUROMOD.*

Union.[6] The scope of these calculations is usually limited to direct taxes and (cash) benefits and current work incentives, though extensions also account for consumption taxes and taking a life-cycle labor supply perspective (see Kotlikoff and Rapson, 2007). Graphically, METRs can be illustrated with a budget constraint chart that plots net income against gross earnings (or hours worked) (see Adam et al., 2006b; Morawski and Myck, 2010), as the slope of this line corresponds to $1 - \text{METR}$ which is the proportion of additional gross earnings retained by the individual.

Participation tax rates (PTR) are conceptually very similar, indicating the effective tax rate on the extensive margin, or the proportion of earnings paid as taxes and lost due to benefit withdrawal if a person moves from inactivity or unemployment to work. METRs and PTRs are typically between 0% and 100%, with higher rates implying weaker incentives to work (more). Because of nonlinearities and complex interactions in the tax-benefit systems, however, individuals facing greater than 100% (or negative) tax rates may also be found. These often expose unintended effects built into the tax-benefit system. More generally, relatively high values indicate situations that can constrain labor supply and trap people at certain income/employment levels. Marginal effective tax rates

[6] In Figure 24.1 and elsewhere we use the official country acronyms for the EU countries. These are (using the official country ordering): Belgium (BE), Bulgaria (BG), Czech Republic (CZ), Denmark (DK), Germany (DE), Estonia (EE), Ireland (IE), Greece (EL), Spain (ES), France (FR), Italy (IT), Cyprus (CY), Latvia (LV), Lithuania (LT), Luxembourg (LU), Hungary (HU), Malta (MT), Netherlands (NL), Austria (AT), Poland (PL), Portugal (PT), Romania (RO), Slovenia (SI), Slovakia (SK), Finland (FI), Sweden (SE), United Kingdom (UK).

and participation tax rates are hence useful indicators to assess whether the tax-benefit system may limit employment for certain individuals. These are also central parameters in assessing optimal tax design. See Immervoll et al. (2007) and Brewer et al. (2010) for empirical applications. Figure 24.1 illustrates how in many countries there is a considerable spread in the value of the METRs even before considering the extremes of the distributions. This demonstrates how an analysis using work incentive indicators based on calculations for average or representative cases may be quite misleading.

Replacement rates (RR) complement participation tax rates, showing the level of out-of-work income relative to in-work disposable income (see, e.g., Immervoll and O'Donoghue, 2004). High replacement rates also reflect low financial incentives to become (or remain) employed. Compared to METRs and PTRs, negative values are even more exceptional (though not ruled out altogether). RRs are often calculated separately for the short-term and long-term unemployed to reflect differences in the level of unemployment benefits depending on unemployment duration. As work incentive indicators, PTRs and RRs are calculated for nonworking persons for whom potential employment income is not observed, and, hence, the latter must be either predicted or assumed.[7]

Although PTRs and RRs both describe work incentives on the extensive margin, they have a different focus and characteristics (Adam et al., 2006a). For instance, if taxes and benefits are changed so that net income increases by the same amount for the out-of-work and in-work situation (e.g., corresponding to a lump-sum transfer), then the replacement rate would typically increase while the participation rate remained unchanged. This is because the tax burden on additional income does not change while, in relative terms, working becomes less attractive. On the other hand, RRs remain constant if out-of-work and in-work net income increase by the same proportion (but for PTRs this is not the case).

Although these three indicators are used to measure work incentives for a particular individual by changing *individual* gross earnings (and labor market status), the effect on disposable income is assessed at the *household* level because this is usually considered to be the more relevant unit of assessment for benefits and unit of aggregation when measuring living standards.[8] Each measure can be also decomposed to show the effect of specific tax-benefit instruments, for example, income taxes, social insurance contributions, and benefits.

24.2.2.2 Indicators of Automatic Stabilization

Another closely related group of indicators characterize how tax-benefit systems act as automatic stabilizers for income or unemployment shocks, as indicated by the extent to which (aggregate) household income or tax revenue fluctuations are moderated

[7] For example, OECD calculates these indicators assuming various income levels in the range of 33–150% of average wage (AW).

[8] In principle other units of aggregation within the household could be specified.

without direct government action. These focus on exogenous shocks rather than individual incentives to alter labor supply. Apart from this, the calculations are technically very similar to the previous group, with the main differences related to interpretation.[9]

Estimates based on microdata go back at least to Pechman (1973), who simulated income tax revenues in the US for 1954–1971 and showed how much tax liabilities change in absolute terms compared to changes in income (at the aggregate level), characterized as built-in flexibility. Although this is very similar to marginal effective tax rates, the interpretation is different and focused on the macro-level and government revenue side rather than at the individual. A closely related measure captures the elasticity of tax liability with respect to changes in incomes, or percentage increase in taxes for a 1% change in income, though as Auerbach and Feenberg (2000) point out, this mainly reflects the progressivity of taxes because it does not capture whether the tax burden is high or low.

More recently, Auerbach and Feenberg (2000) estimate the aggregate change in taxes when increasing all (taxable) income (and deductions) for each individual by 1% to measure the responsiveness of tax revenues to income changes for the US. They find that, over the period 1962–1995, income taxes offset between 18% and 28% of variation in before-tax income (at the aggregate level). Similarly, Mabbett and Schelke (2007) simulate a 10% increase in individual earnings for 14 EU countries and estimate both the responsiveness (i.e., elasticity) of various tax-benefit instruments and the overall stabilization effect of the system. According to their estimates, the latter varies from 31% in Spain to 57% in Denmark.

Dolls et al. (2012) model a *negative* income shock in which household gross incomes fall by 5% and an unemployment shock with household income at the aggregate level decreasing also by 5%, covering both the US and a large number of the EU countries. Although the proportional income shock is distribution-neutral, the unemployment shock is asymmetric because not all households are affected. They find that tax-benefit systems absorb a greater proportion of income variation in the EU compared to the US— 38% (EU) versus 32% (US) of the income shock and 47% (EU) versus 34% (US) of the unemployment shock (see Figure 24.2). This difference is largely explained by the higher coverage and generosity of unemployment benefits in Europe. Automatic stabilizers in the case of an unemployment shock are basically replacement rates for a transition from employment to nonwork at the aggregate level. Rather than work incentives (as discussed in the previous section), they reflect how much the tax-benefit system absorbs (market) income losses due to becoming unemployed or exiting the labor market altogether.

Instead of focusing on aggregate stabilization, Fernández Salgado et al. (2014) analyze the distribution of replacement rates when simulating the unemployment shock in six EU

[9] One technical nuance concerns the treatment of multiperson households. Although work incentives are typically estimated for each household member separately, holding earnings of other household members constant, in the case of automatic stabilizers, changes are simulated for all of the relevant population at once.

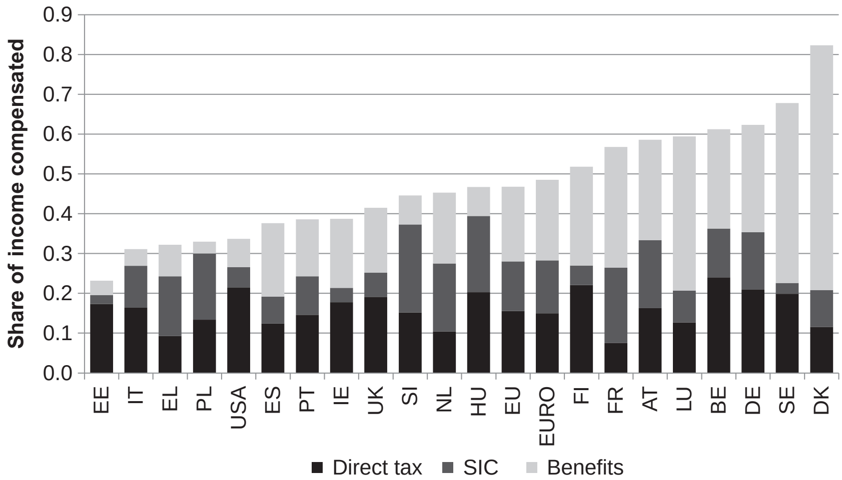


Figure 24.2 Share of income compensated by the tax benefit system in the case of an unemployment shock. *Notes: The unemployment shock corresponds to an increase in the unemployment rate such that the total household income decreases by 5%. Countries are ranked by the share of income compensated. EU and EURO are the population-weighted averages of 19 EU and 13 eurozone countries, respectively, included here. Estonia joined the eurozone later and is here excluded from that group. Source: Dolls et al. (2012), using EUROMOD and TAXSIM.*

countries due to the Great Recession. They distinguish between short- and long-term unemployment, and their findings confirm higher replacement rates in the short term and point to serious challenges for minimum income schemes to cope with the consequences of this crisis in the longer term. They also highlight the important role of incomes of other household members in boosting replacement rates.

24.2.2.3 Indicators of Household Composition Effects

Another type of indicator based on microsimulation captures the effect of changes in household sociodemographic characteristics in order to identify the marginal effect of the tax-benefit system due to particular household configurations. For example, Figari et al. (2011b) apply this approach to calculate "child contingent" incomes estimated as the change in household disposable income for families with children as if they did not have children. They argue that "child-contingent" incomes, capturing not only transfers net of taxes but also tax concessions, account more precisely for the full net support provided through tax-benefit systems to families with children, as compared to simply considering (gross) benefit payments labeled explicitly for children or families, as is typically the case using the information directly available from the survey data. As shown in Figure 24.3, the net value can be greater than the gross if there are tax concessions or child supplements in benefits labeled for other purposes, and the gross value can be greater than the net if the benefits are taxed or reduced because of other interactions.

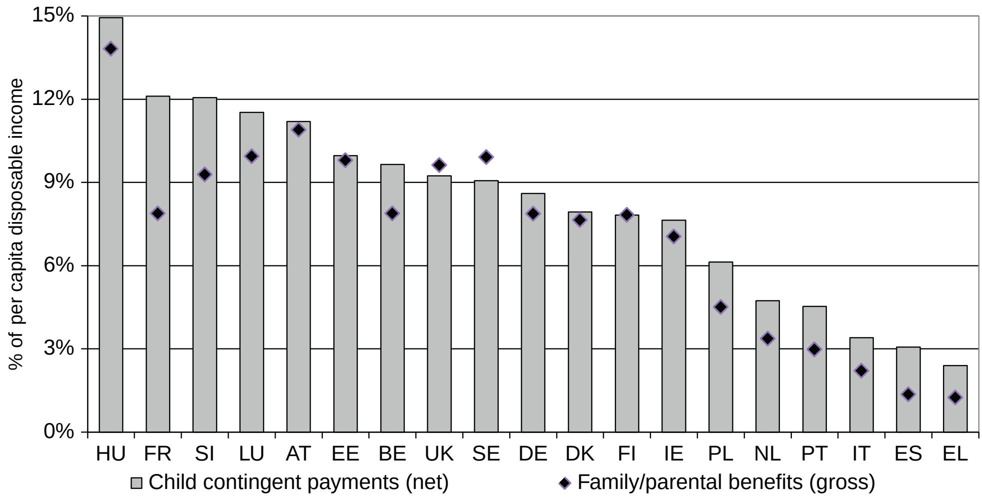


Figure 24.3 Total net child-contingent payments versus gross family/parental benefits per child as a percentage of per capita disposable income. *Notes: Countries are ranked by total net child-contingent payments. Source: Figari et al. (2011b), using EUROMOD.*

24.3. THE EFFECTS OF POLICY CHANGES ON INCOME DISTRIBUTION

24.3.1 A Basic Example

The simplest use of a tax-benefit microsimulation model involves calculating the effects of a policy change on household income, without changing any of the characteristics of the household members. An example might be an increase in the amount of an existing universal child benefit. The model would take account of the increase in payment per eligible child, any clawback through the system of means-tested benefits (if the child benefit is included in the income assessment for these benefits), any clawback if the benefit is taxed or included in the base for contributions, and any other relevant interaction with the rest of the tax-benefit system. Even a simple reform involves quite complicated arithmetic and ignoring the interactions would give misleading results. This is illustrated in Appendix A with a concrete example comparing the effects of doubling the UK child benefit at two points in time: 2001 and 2013. Although the structure of child benefit itself has remained the same, the net effect of changes to it is quite different because of changes to the interactions with the rest of the tax-benefit system. The interactions matter and need to be accounted for in understanding the effects of policy changes and in designing policy reforms.

The financing of such a reform would also need to be considered. For example, if the net cost was met by a percentage point increase in all rates of income tax, this increase might also have knock-on effects (e.g., if the assessment of any means-tested benefits depended on after-tax income), and then iterations of the model would be needed to find a revenue-neutral solution to the tax rate increase. The "revenue neutral" package could then be evaluated relative to the prereform situation, in terms of its effect on the income distribution and an analysis of gainers and losers.

Of course, in the new situation some people affected will wish to change their behavior in response to the change in some way and at some point in time. One might expect labor supply and fertility to be affected and, depending on the specifics of the system and the change in it, so might other dimensions of behavior. As Bourguignon and Spadaro (2006) point out, it is important to be clear about when these second order effects can, and cannot, be neglected. We return to this issue in Section 24.3.3.

In any case, an "overnight" or "morning after" analysis, as the pure arithmetic effect is often called, is clearly of value in its own right as the immediate effect might be relevant to a particular research question. Moreover, the mechanics of the way in which policy reforms impact on incomes are relevant for improving design, and it will often be important to identify how much of the overall effect on income can be attributed to the direct effect.

24.3.2 Formal Framework

24.3.2.1 Decomposing Static Policy Effects

Tax-benefit models provide information on the distribution of household disposable income and its components under various policy scenarios, allowing the effects of policies to be inferred from a comparison of different scenarios. As such, the application of the microsimulation method starts by defining an appropriate baseline and a counterfactual scenario. The latter corresponds to the state *after* policy changes (i.e., how the world would look after implementing new policies) in forward-looking analysis or the state *before* policy changes (i.e., how the world would have looked without new policies or what would happen if policy changes where rolled back) in the case of backward-looking analysis.

Drawing on Bargain and Callan (2010) and Bargain (2012a), we provide a formal framework for decomposing changes in household income to separate the effects of policy changes.[10] Mathematical formulation helps to avoid ambiguities about how exactly a counterfactual scenario is defined, which often arise in empirical microsimulation applications relying only on textual descriptions. Furthermore, full decomposition (rather than only focusing on the role of policy changes) has clear advantages by drawing attention to the fact that the (marginal) contribution of a given component is evaluated conditional on the values of other components, and, hence, the overall change in income can be decomposed in multiple ways. Decomposing all components (at once) also helps to ensure that

[10] There is a notable strand in the economic literature focusing on the decomposition of income distributions, reviewed recently in Fortin et al. (2011). This, however, is primarily concerned with wage distributions, ignoring the role of tax-benefit policies.

these are consistently derived. Apart from small technical modifications,[11] we closely follow Bargain and Callan's original approach but broaden its scope by showing that a wider range of applications can be interpreted within the same framework.

Let us denote household sociodemographic (and labor market) characteristics with a vector c and household original income[12] (i.e., income before adding cash benefits and deducting direct taxes) with a vector x. The net transfer via the tax-benefit system k (i.e., total cash benefit entitlement less total direct tax liability) for a household with characteristics c and income x is denoted as a function $f_k(c, x, m_k)$, where, following Bargain and Callan (2010), we distinguish between the structure of the tax-benefit system f_k and the various monetary parameters m_k it takes as arguments (e.g., tax brackets, benefit amounts). $f_k(c, x, m_k)$ is positive if public pensions and cash benefits received by a given household exceed direct taxes for which the household is liable, and it is negative if the opposite holds. Household disposable income y is then

$$y_k(c, x, m_k) = x + f_k(c, x, m_k). \tag{24.1}$$

In the simplest case, where original income and household characteristics can be assumed to remain constant, the effect of policy changes ($A \to B$) on disposable income is

$$\Delta y = y_B(c, x, m_B) - y_A(c, x, m_A). \tag{24.2}$$

This corresponds to how the effects of proposed or hypothetical tax-benefit reforms are typically studied, as "morning-after" changes with the policy rules before and after referring (implicitly) to the same time period. There are numerous examples of such exercises (e.g., Callan and Sutherland, 1997; Figari, 2010; Matsaganis and Flevotomou, 2008; Matsaganis et al., 2006; Paulus and Peichl, 2009).

Next, let us consider the case of analyzing the effect of policy changes over time. Accounting explicitly for the time span over which policy changes are considered introduces additional complexities for defining an alternative scenario. It is important to ensure that the baseline and the counterfactual refer to the same time period, and if there is a time gap between the existing policies and the counterfactual, then one or the other must be adjusted to reflect that. For example, when analyzing the effect of policy change in $t+1$, it may not be sufficient to assume that the alternative would have been simply period t policies continuing (in nominal terms) in period $t+1$, even though this is often

[11] We carry out decomposition in steps starting from the policy effects—our key interest—and other effects, then introducing further splits. Such nesting helps to ensure consistency between the various components and across different combinations. We also distinguish sociodemographic characteristics (c) from original incomes (x).

[12] This includes market incomes from employment, self-employment, property and investments, and other nonpublic income sources such as private pensions and transfers between households. It is also known as "pretax and pretransfer income" in the literature.

implicitly done. One should consider how existing policies in nominal terms *would* have evolved otherwise, given the legal rules or usual practice of indexation of policy parameters, or *should* have evolved. The importance of the time factor becomes even more obvious when considering policy changes over a longer period. We will return later to the question of what is an appropriate basis for indexing monetary parameters in the counterfactual scenario, but for now we simply denote such a factor as p.

First, the total change in disposable income for a given household can be decomposed to show first order policy effects (or mechanical effects) *conditional* on household characteristics and original incomes in the end-period B (denoting the start-period with A):

$$\begin{aligned}\Delta y &= y_B(c_B, x_B, m_B) - y_A(c_A, x_A, m_A) \quad \text{(total change)}\\ &= y_B(c_B, x_B, m_B) - y_A(c_B, x_B, pm_A) \quad \text{(policy effect)}\\ &\quad + y_A(c_B, x_B, pm_A) - y_A(c_A, x_A, m_A) \quad \text{(other effect)}\\ &= \Delta y_I^P + \Delta y_I^O\end{aligned} \tag{24.3}$$

Here we are implicitly assuming that we are dealing with panel data, with characteristics and original income for the same household being observed in several periods. The total change for the same household cannot be observed with multiple waves of cross-sectional datasets; however, as explained further below, the same decomposition approach can be also applied at the group-level (e.g., the bottom decile group) or to statistics summarizing the whole income distribution (such as various inequality indices). Importantly, household characteristics are only required at a single point in time to calculate the policy effect (in absolute terms).

Noting the symmetry of the decomposition, the other effect can be decomposed further into two subcomponents separating the impact of change in household characteristics and nominal levels. The effect due to changes in characteristics can be measured either in end-period income levels:

$$\begin{aligned}\Delta y_I^O = &\ y_A(c_B, x_B, pm_A) - y_A(c_A, px_A, pm_A) \quad \text{(change in characteristics)}\\ &+ y_A(c_A, px_A, pm_A) - y_A(c_A, x_A, m_A) \quad \text{(change in nominal levels)}\end{aligned} \tag{24.4}$$

or start-period incomes

$$\begin{aligned}\Delta y_I^O = &\ y_A(c_B, x_B, pm_A) - y_A(c_B, p^{-1}x_B, m_A) \quad \text{(change in nominal levels)}\\ &+ y_A(c_B, p^{-1}x_B, m_A) - y_A(c_A, x_A, m_A) \quad \text{(change in characteristics)}\end{aligned} \tag{24.5}$$

The term capturing the effect of "change in nominal levels" measures how household disposable income is affected if original income and all money-metric policy parameters change in the same proportion. As Bargain and Callan (2010) pointed out, tax-benefit systems are typically homogenous of degree 1, meaning that in such a case, household disposable income would also change by the same factor:

$$py(c, x, m) = y(c, px, pm) \tag{24.6}$$

They illustrate this with a hypothetical example involving a basic income and a flat tax and find empirical support for Ireland and France.[13]

In principle, the term reflecting the impact of changes in characteristics could be split further, distinguishing between changes in sociodemographic (and labor market) characteristics c and movements in original incomes x. Again, there would be two possible combinations that could be obtained by introducing a new term—either $y_A(c_B, px_A, pm_A)$ or $y_A(c_A, x_B, pm_A)$ with Equation (24.4).

Second, the change in disposable income can be decomposed to assess policy effects conditional on household characteristics and original incomes in the start-period A:

$$\begin{aligned} \Delta y &= y_B(c_B, x_B, m_B) - y_B(c_A, x_A, p^{-1}m_B) \quad \text{(other effect)} \\ &\quad + y_B(c_A, x_A, p^{-1}m_B) - y_A(c_A, x_A, m_A) \quad \text{(policy effect)} \\ &= \Delta y_{II}^O + \Delta y_{II}^P \end{aligned} \tag{24.7}$$

The other effect can be now decomposed again such that the effects due to changes in characteristics are measured in end-period incomes:

$$\begin{aligned} \Delta y_{II}^O &= y_B(c_B, x_B, m_B) - y_B(c_A, px_A, m_B) \quad \text{(change in characteristics)} \\ &\quad + y_B(c_A, px_A, m_B) - y_B(c_A, x_A, p^{-1}m_B) \quad \text{(change in nominal levels)} \end{aligned} \tag{24.8}$$

or start-period incomes

$$\begin{aligned} \Delta y_{II}^O &= y_B(c_B, x_B, m_B) - y_B(c_B, p^{-1}x_B, p^{-1}m_B) \quad \text{(change in nominal levels)} \\ &\quad + y_B(c_B, p^{-1}x_B, p^{-1}m_B) - y_B(c_A, x_A, p^{-1}m_B) \quad \text{(change in characteristics)} \end{aligned} \tag{24.9}$$

Altogether there are four ways to decompose the overall change in income, given the initial split into the direct policy effect and the other effect.

Until now, we have focused on a single household, though it is straightforward to derive the aggregate change in disposable income by summing income differences (and subcomponents) across all households:

$$\Delta Y = \sum_i \Delta y^i \tag{24.10}$$

Decomposition can also be applied to any distributional statistic D calculated for a specific subgroup, as with average income among households with elderly, or summarizing the whole income distribution, ($\boldsymbol{y}$), as with the Gini coefficient or the headcount poverty ratio. For example, Equation (24.3) would then become (indicating vectors in bold):

[13] Bargain and Callan (2010) refer to the personal income tax in Germany as one of the few examples for which this property does not hold, due to its unique quadratic functional form.

$$\begin{aligned}\Delta D(y) &= D[y_B(\boldsymbol{c}_B, \boldsymbol{x}_B, \boldsymbol{m}_B)] - D[y_A(\boldsymbol{c}_A, \boldsymbol{x}_A, \boldsymbol{m}_A)] \quad \text{(total change)} \\ &= D[y_B(\boldsymbol{c}_B, \boldsymbol{x}_B, \boldsymbol{m}_B)] - D[y_A(\boldsymbol{c}_B, \boldsymbol{x}_B, p\boldsymbol{m}_A)] \quad \text{(policy effect)} \\ &\quad + D[y_A(\boldsymbol{c}_B, \boldsymbol{x}_B, p\boldsymbol{m}_A)] - D[y_A(\boldsymbol{c}_A, \boldsymbol{x}_A, \boldsymbol{m}_A)] \quad \text{(other effect)}\end{aligned} \tag{24.11}$$

In the case of scale-invariant distributional measures (see Cowell, 2000) and linearly homogenous tax-benefit systems, the decomposition of other effects (Equations 24.4, 24.5, 24.8, 24.9) simplifies because the effect of a change in nominal levels becomes (approximately) zero at the population level.[14] Furthermore, Equation (24.4) is now equivalent to Equation (24.5), and Equation (24.8) is equivalent to Equation (24.9), reducing the overall number of combinations from four to two.

We now return to what would be an appropriate basis for choosing the indexation factor p. Bargain and Callan (2010) have argued for using the growth of average original incomes, expressed as $p = \overline{x}_B/\overline{x}_A$, to obtain a "distributionally neutral" benchmark. This would broadly ensure that aggregate disposable income rises (or falls) in proportion to an increase (or a decrease) in aggregate original incomes; in other words, the overall tax burden and expenditure level remain constant in relative terms.[15] Nevertheless, disposable income for a given household could still grow at a higher (or lower) rate than their original income if the latter grows less (or more) than on average. However, there are alternatives ways of choosing p, depending on the chosen conception of "neutrality." For example, basing it on a consumer price index would be appropriate if the point was to ensure a constant absolute standard of living (on average). Clark and Leicester (2004) contrast price-indexation with indexation based on the nominal GDP, and they show that the choice matters for results. There is no clear consensus in the literature on decomposition regarding the most appropriate choice of index.

Finally, there is the issue of how to deal with path dependency and multiple combinations for decomposition. Can some combinations be preferred over others or can different combinations somehow be brought together? In some cases, one might be limited to specific combinations by data constraints. The prime example here is ex ante analysis of (implemented) policy changes before microdata of actual postreform incomes become available (e.g., Avram et al., 2013). Relying on estimates for p, one could already quantify the effect of policy changes (with Equation 24.7), but both start- and end-year datasets are needed to assess other effects. Given that there are no clear arguments for preferring one particular combination over another, all variants should be covered. Bargain and Callan (2010) adopt the Shorrocks–Shapley approach (Shorrocks, 2013) to summarize various combinations, essentially averaging the effect of a given component

[14] Notice that, with each subcomponent aggregated separately (rather than aggregating the differences between various components), it is possible to carry out such exercise on two waves of cross-sectional data without necessarily using panel data.

[15] This holds for linearly homogenous tax-benefit systems, but nonlinear elements make this an approximation. Nevertheless, as demonstrated in Callan et al. (2007) for Ireland, the bias would typically be small.

across all combinations. In this way, results conditional on household characteristics in each period are given equal weights.

Other examples where such decomposition has been used explicitly include Bargain (2012b), Bargain et al. (2013b), and Creedy and Hérault (2011). In addition, there is a large literature that documents similar assessments within less formal frameworks (for example, see Clark and Leicester, 2004; Thoresen, 2004).

24.3.2.2 Specific Applications

24.3.2.2.1 Actual Versus Counterfactual Indexation of Policy Parameters

Any system which is not fully indexed with respect to growth in (average) private incomes or prices would result in the erosion of the relative value of benefit payments and increased tax burden through so-called bracket creep (or fiscal drag). Furthermore, it is essential to acknowledge that keeping a tax-benefit system unchanged also impacts household incomes (unless the distribution of household original income is also constant over time). Let us consider the change in household disposable income in such a case using our notation from above:

$$\Delta y = y(c, x_B, m) - y(c, x_A, m) \tag{24.12}$$

Following the decomposition framework in the more general case above, Equation (24.12) can be split again into three terms: the policy effect, changes in original incomes, and the change in nominal levels (i.e., change in disposable income if both original incomes and policy parameters were scaled up by the same factor). The policy effect would now reflect the outcome of keeping policy parameters constant in nominal terms and can be calculated as $y(c, x_B, m) - y(c, x_B, pm)$ or $y(c, x_A, p^{-1}m) - y(c, x_A, m)$. In a typical case, p is positive, reflecting growth in private incomes (or consumer prices), and, hence, the policy effect would be negative (i.e., income-reducing). This is because a positive p implies higher benefit amounts and tax bands in the counterfactual scenario and translates into higher disposable incomes (for the same original incomes) compared to disposable incomes under tax-benefit rules when these are kept nominally constant. This has been studied, for example, by Immervoll (2005), Immervoll et al. (2006b) and Sutherland et al. (2008). It is also important to realize that if p is negative, meaning average original incomes (or prices) fall, and a tax-benefit system is kept nominally constant, then households' tax burdens fall in relative terms.

24.3.2.2.2 Policy Swaps

An analogous type of exercise to that comparing the effects of policies across time in one country involves assessing the effects of policies from one country (A) when simulated in another (B), the so-called policy swaps. The starting point is again Equation (24.3), but instead of comparing the effects of two different national policy regimes on the same population and distribution of original incomes, the aim is to compare the effects of a particular set of "borrowed" policies on different populations and income distributions. Some studies focus on the effects of several alternative systems in one particular country (one-way swaps), and others carry out two-way swaps sometimes involving more than two countries in a series of swaps. Section 24.3.6 discusses some examples of such studies. Instead of growth in income over time and the relative movement in tax-and benefit parameters, the nature of p has to do with difference in nominal levels of original income across countries. Often there are additional complexities involved in maintaining correspondence with original policies, especially if more than one pair-wise comparison is made. Attempts so far have aimed to keep the values of parameters fixed in relative terms, for example, in connection to average income or in order to maintain budget neutrality.

24.3.2.2.3 Microsimulation-Based Indicators

The same framework can be used to describe microsimulation-based indicators, designed to capture some inherent characteristics of a given tax-benefit system, which are not directly observable. The nature of these was already explained in Section 24.2.2, and here we formalize the key definitions. Overall, these indicators show how household disposable income reacts to changes in people's gross earnings and circumstances (for a given tax-benefit system):

$$\Delta y = y(c_B, x_B, m) - y(c_A, x_A, m) \tag{24.13}$$

Using our notation, we can express marginal effective tax rates (METR) as follows:

$$\mathrm{METR} = 1 - [y(c, x + d, m) - y(c, x, m)]/d \tag{24.14}$$

where the change in household disposable income is divided by the margin (d) used to increment gross earnings (x) of a given household member, yielding a relative measure. This is further deducted from one to show the part of additional earnings which is taxed away.

In the case of participation tax rates (PTR), both earnings (x) and other household characteristics (c) are adjusted to reflect the change in labor market status, as with the change from inactivity or unemployment (A) to work (B):

$$\mathrm{PTR} = 1 - [y(c_B, x_B, m) - y(c_A, x_A, m)]/(x_B - x_A) \tag{24.15}$$

The relative income change is again deducted from one to reflect the effective tax rate at this margin. (Note that this could be further simplified as $x_A = 0$.) Replacement rates (RR) are simply calculated as the ratio of out-of-work disposable income (A) to in-work disposable income (B):

$$\mathrm{RR} = y(c_A, x_A, m)/y(c_B, x_B, m) \tag{24.16}$$

Finally, indicators based on counterfactuals reflecting only changes in household socio-demographic characteristics (c) can be calculated as $\Delta y = y(c_A, x, m) - y(c_B, x, m)$.

For example, "child-contingent" incomes would show the change in household disposable income for families with children (A) compared to the income if they did not have children (B).

24.3.2.3 Decomposition With Labor Supply Changes

So far we have focused on the static effects of policy changes, whereby potential behavioral reactions have been absorbed by the component capturing changes in household characteristics more generally. Following Bargain (2012a), we now extend the previous case and explicitly account for behavioral changes in the form of labor supply adjustments due to policy changes. For this purpose, we slightly change the notation from x_k to x_k^l, which refers to original incomes by population with characteristics c_k based on labor supply choices made under the policy system l. (As such, the meaning of x_k^k is exactly the same as x_k before and, hence, will be shortened to the latter.) This allows the term "changes in characteristics" to be further split into two components—labor supply adjustments following changes in policy rules ($A \to B$) and other effects due to changes in the population structure c (assumed to be exogenous to tax-benefit policy changes, at least in the short and the medium term). We can now express the overall change in household disposable income as a sum of four components: direct (or mechanical) policy effect, labor supply reactions, change in nominal levels, and change in characteristics.

Decomposing Equation (24.4) and combining it with Equation (24.3), we can separate the behavioral effects comparing disposable income with labor supply under the initial and the new policy rules, expressed in terms of initial policy rules (y_{A}, pm_{A}) and either start-period household characteristics c_A:

$$\begin{aligned}\Delta y = & \; y_B(c_B, x_B, m_B) - y_A(c_B, x_B, pm_A) \quad \text{(direct policy effect)} \\ & + y_A(c_B, x_B, pm_A) - y_A\left(c_A, px_A^B, pm_A\right) \quad \text{(change in characteristics)} \\ & + y_A\left(c_A, px_A^B, pm_A\right) - y_A(c_A, px_A, pm_A) \quad \text{(behavioral effects)} \\ & + y_A(c_A, px_A, pm_A) - y_A(c_A, x_A, m_A) \quad \text{(change in nominal levels)}\end{aligned} \tag{24.17}$$

or end-period household characteristics c_B

$$\begin{aligned}\Delta y = & \; y_B(c_B, x_B, m_B) - y_A(c_B, x_B, pm_A) \quad \text{(direct policy effect)} \\ & + y_A(c_B, x_B, pm_A) - y_A\left(c_B, x_B^A, pm_A\right) \quad \text{(behavioral effects)} \\ & + y_A\left(c_B, x_B^A, pm_A\right) - y_A(c_A, px_A, pm_A) \quad \text{(change in characteristics)} \\ & + y_A(c_A, px_A, pm_A) - y_A(c_A, x_A, m_A) \quad \text{(change in nominal levels)}\end{aligned} \tag{24.18}$$

Decomposing Equation (24.8) instead and combining it with Equation (24.7) allow the behavioral effects to be expressed in terms of new policy rules (y_B, m_B) and, again, start-period household characteristics c_A:

$$\begin{aligned}\Delta y = {} & y_B(c_B, x_B, m_B) - y_B\left(c_A, px_A^B, m_B\right) \quad \text{(change in characteristics)}\\ & + y_B\left(c_A, px_A^B, m_B\right) - y_B(c_A, px_A, m_B) \quad \text{(behavioral effects)}\\ & + y_B(c_A, px_A, m_B) - y_B(c_A, x_A, p^{-1}m_B) \quad \text{(change in nominal levels)}\\ & + y_B(c_A, x_A, p^{-1}m_B) - y_A(c_A, x_A, m_A) \quad \text{(direct policy effect)}\end{aligned} \tag{24.19}$$

or end-period household characteristics c_B[16]:

$$\begin{aligned}\Delta y = {} & y_B(c_B, x_B, m_B) - y_B\left(c_B, x_B^A, m_B\right) \quad \text{(behavioral effects)}\\ & + y_B\left(c_B, x_B^A, m_B\right) - y_B(c_A, px_A, m_B) \quad \text{(change in characteristics)}\\ & + y_B(c_A, px_A, m_B) - y_B(c_A, x_A, p^{-1}m_B) \quad \text{(change in nominal levels)}\\ & + y_B(c_A, x_A, p^{-1}m_B) - y_A(c_A, x_A, m_A) \quad \text{(direct policy effect)}\end{aligned} \tag{24.20}$$

Modeling behavior and in particular labor supply is discussed in more detail in the next section.

24.3.3 Modeling Behavioral Changes

24.3.3.1 Accounting for Individual Reactions

The impact of policies on individual behavior, through incentives and constraints, is at the core of economics, and behavioral microsimulation models are valuable tools for providing insights into the potential behavioral reactions to changes in the tax-benefit system and, consequently, on their effect on economic efficiency, income distribution, and individual welfare (Creedy and Duncan, 2002). Nevertheless, it is important to be clear when the second-order effects can, and cannot, be neglected. To capture the individual effect of reforms, it is not always necessary to quantify behavioral responses, on the assumption that the effects of the policy changes are marginal to the budget constraint (Bourguignon and Spadaro, 2006).

Of course, it is not possible to judge a priori whether the behavioral response is large or ignorable. Judgments must necessarily be made on an *ad hoc* basis, using available evidence and related to the context of the analysis and how results are to be interpreted. If behavior is known to be constrained (e.g., in the case of labor supply adjustments at times of high unemployment), then behavioral responses might be ignored, and it is sufficient to consider the results of the analysis in terms of changes in income (rather than welfare). If static indicators of work incentives, such as marginal effective tax rates and participation tax rates, change very little as a result of a policy change, then one can assume that labor supply responses driven by substitution effects will be small. If the change in income with and without modeled behavioral response is expected to be rather similar, then given the error in the modeling of behavior and in the static microsimulation estimates themselves,

[16] We skip decompositions based on Equations (24.5) and (24.9) because the terms for behavioral effects and change in characteristics are simply scaled down by p (if the homogeneity property holds), compared to Equations (24.4) and (24.8), respectively.

going to the trouble of modeling responses may not be worthwhile (Pudney and Sutherland, 1996).

Moreover, being clear about the relevant time period is important. From a policy-making viewpoint, it is the effect on the income distribution and on the public budget in the year of the reform that often matters. Most tax and benefit policy changes are made year to year and are fine-tuned later if necessary. On the one hand, behavior takes time to adapt to changing policies, partly because of constraints, adjustment costs, and lack of information or understanding. This applies most obviously to fertility but also to labor supply in systems where full information about the policy rules is not available until the end of the year (after labor supply decisions have already been acted on). On the other hand, changes in behavior may also happen in anticipation of the policy being implemented, with short-term responses larger than long-terms effects. This may apply particularly to tax planning behavior and is well-illustrated by the case of an announcement in the UK in 2009 of a large increase in the top rate of tax on high incomes in the 2010–2011 tax year. Major forestalling of income by those who would pay the additional tax and were in a position to manipulate the timing of their income resulted in an unexpected increase in tax revenue in the 2009–2010 tax year and a corresponding reduction in the following year (HMRC, 2012).

In some situations the morning-after effect is the most relevant when considering short-term policy adjustments and equilibrium (or partial equilibrium) considerations are not particularly relevant. Furthermore, if indicative results are needed quickly because reform is imperative, then in the absence of an already-estimated and tested behavioral model, static results with the appropriate "health warnings" are still more informative than nothing at all.

Nevertheless, it is widely recognized that, depending on the policy change being analyzed, ignoring behavioral reactions can lead to misleading estimates of the impact of the policy reform on the income distribution and the macroeconomic consequences (Bourguignon and Spadaro, 2006), as is also illustrated by the tax-planning example.

At the other extreme, modeling behavioral responses in the case of very large changes to policy poses challenges for the empirical basis of behavioral modeling. For example, replacing an existing tax-benefit system with a combination of a basic income and a flat tax such that no income fell below the poverty threshold would presumably result in large changes in many dimensions of behavior which are unlikely to be correctly captured by the labor supply models that are used traditionally.

Despite the long tradition of modeling behavior in economics, the behavioral reactions to changes in the tax system that are most commonly analyzed are related to labor supply (starting with the seminal contributions of Aaberge et al., 1995 and van Soest, 1995) and program participation (Keane and Moffitt, 1998), feeding into a growing literature, which is characterized by an increasing level of econometric sophistication. The same level of development does not yet apply to other research areas in which microsimulation models have been used, such as investigating the potential effects of tax policies on consumption (Creedy, 1999b; Decoster et al., 2010), saving (Boadway and Wildasin, 1995; Feldstein and Feenberg, 1983), and housing (King, 1983), at least partly due to a lack of suitable data.

24.3.3.2 Labor Supply Models

There is a general consensus in the literature about using (static) discrete choice models to simulate the individual labor supply reactions to changes in the tax-benefit system.[17] Such models are structural because they provide direct estimations of preferences over income and hours of work, through the specification of the functional form of the utility function. Discrete choice models belong to the family of random utility maximization models (McFadden, 1974) that allow the utility function to have a random component (usually following the extreme value distribution), affecting the optimal alternative in terms of utility level associated with each choice.

The discrete choice character of the models is due to the assumption that utility-maximizing individuals and couples choose from a relatively small number of working hour combinations, which form the personal choice set. Each point in the choice set corresponds to a certain level of disposable income given the gross earnings of each individual (derived using the observed or predicted wage), other incomes, and the tax-benefit system rules simulated by means of a tax-benefit microsimulation model taking into account the sociodemographic characteristics of the family. The nonlinear and nonconvex budget sets determined by complex tax-benefit systems provide a primary source of identification of the model itself. Most of the discrete choice models based on the van Soest (1995) approach assume that the same choice sets are defined and available for each individual and that an individual has the same gross hourly wage for each such alternative. Ilmakunnas and Pudney (1990) is one of the few exceptions in the literature, allowing the hourly wage to be different according to the number of hours offered by each individual. Aaberge et al. (1995) provide a more flexible specification that defines the alternatives faced by the individuals in terms of a set of a wage rate, hours of work, and other job-related characteristics. The wage rate can differ for the same individual across alternatives, the hours of work are sampled from the observed distribution, and the availability of jobs of different types can depend on individual and institutional characteristics.

Regardless of the econometric specifications, the sample is usually restricted to individuals who are considered "labor supply-flexible" in order to exclude individuals whose labor choices are affected by factors that are not or cannot be controlled for in the labor supply model. Examples of these factors include disability status, educational choices, early retirement, and self-employment status. This represents a limit in the use of the

[17] See Creedy and Kalb (2005) for an extensive review of modeling strategies.

estimated labor supply responses to analyze changes in overall income distribution because, for the individuals not covered by the labor supply models, the behavior is assumed to be inelastic. In most applications, working age individuals within the family are allowed to vary their labor supply independently of each other, and the utility maximization takes place at the family level, considering the income of both partners subject to a pooled income constraint, in line with the unitary model of household behavior. Blundell et al. (2007) provides an example of the structural model of labor supply in a collective setting, excluding the effects of taxes.

Figure 24.4 depicts the main components of a standard labor supply model that uses a static tax-benefit algorithm to generate input for the labor supply estimation and to evaluate the labor supply reactions to policy reforms.

In the prereform scenario (left panel of Figure 24.4), the labor supply model is estimated on the budget set providing a direct estimate of the preferences over income and hours. In the postreform scenario (right panel of Figure 24.4), a new budget set for each family is derived by the tax-benefit model applying the new tax-benefit rules following the simulated reform. Assuming that individual random preference heterogeneity and observable preferences do not vary over time, labor supply estimates from the prereform scenario are used to predict the labor supply effects and the second-round redistributive effects (i.e., when labor supply reactions are taken into account) of the simulated policy reforms. Such effects might come out of an iterative procedure calibrating the policy parameters to ensure revenue neutrality once the labor supply reactions and their effects on tax revenue and benefit expenditure are taken into account.

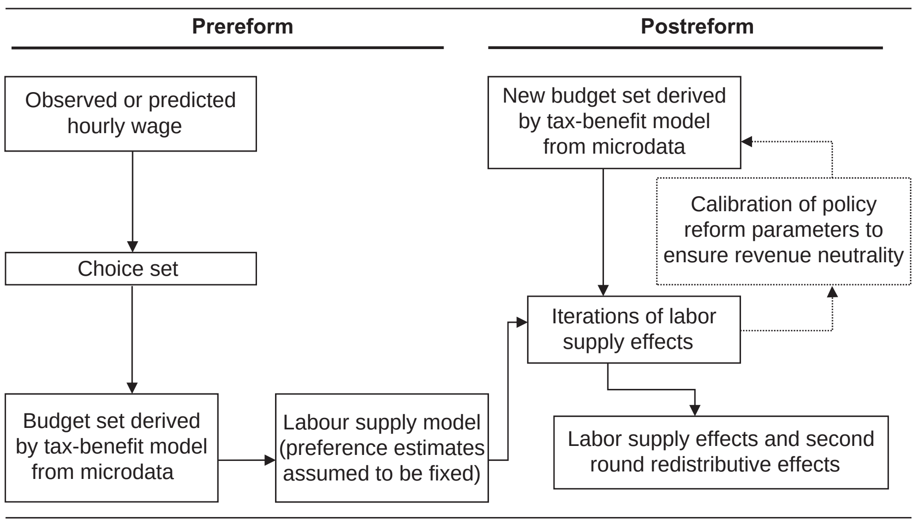


Figure 24.4 Behavioral tax-benefit model and underlying data.

Applications of discrete choice behavioral models are too numerous to be surveyed in this context. Along with many applications focused on the potential effects of specific tax-benefit policies (among others, see Brewer et al. (2009) for a review of analysis of the effects of in-work benefits across countries), labor supply models based on microsimulation models provide labor supply elasticities that can be used in other tax policy research (e.g., Immervoll et al., 2007). Using EUROMOD and TAXSIM, Bargain et al. (2014) provide the first large-scale international comparison of labor supply elasticities including 17 EU countries and the US. The use of a harmonized approach provides results that are more robust to possible measurement differences that would otherwise arise from the use of different data, microsimulation models and methodological choices. Figure 24.5 shows the estimated own-wage elasticities for single individuals and individuals in couples, which suggest substantial scope for the potential impact of tax-benefit reforms on labor supply and hence income distribution, though the differences across countries are found to be smaller with respect to those in previous studies. Bargain et al. (2014) also show the extent to which labor supply elasticities vary with income level which has important implications for the analysis of the equity-efficiency trade off inherent in tax-benefit reforms. To this aim, labor supply models can be used to implement a computational approach to the optimal taxation problem, allowing the empirical identification of the optimal income tax rules according to various social welfare criteria under the constraint of revenue neutrality (Aaberge and Colombino, 2013).

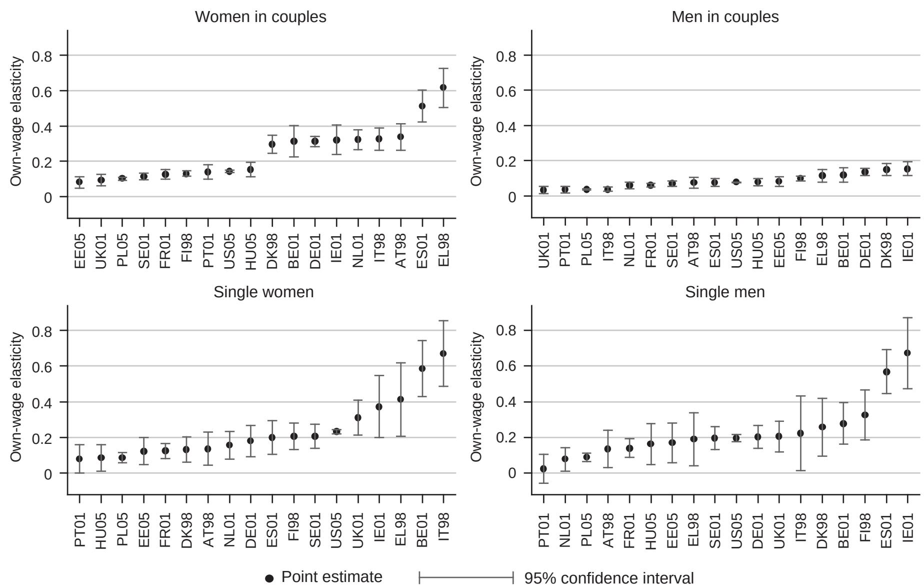


Figure 24.5 Europe and United States: Own-wage elasticities. Source: *Bargain et al. (2014), using EUROMOD and TAXSIM.*

The rapid dissemination of labor supply models, no longer restricted to the academic sphere and increasingly used to inform the policy debate, has been accompanied by a continuing refinement of the econometric specifications. Nevertheless, further improvements are still necessary to model the labor market equilibrium that can emerge as a consequence of a policy simulation (Colombino, 2013), to take into account demand side constraints (Peichl and Siegloch, 2012) and to exploit the longitudinal dimension of micro-data, where this is available, in order to avoid labor supply estimates being potentially biased by individual unobserved characteristics and to consider the state dependence in the labor supply behavior (Haan, 2010).

24.3.4 Macroeconomic Effects

In a basic application of a static microsimulation model, labor market conditions and levels of market income are taken from the underlying data without further adjustments. However, these conditions may change due to policy, and economic and institutional changes, and in order to assess the social consequences of macroeconomic changes, it is important to consider the interactions of the tax-benefit system with new conditions in the labor market and with other macroeconomic effects in general. On the one hand, micro-oriented policies can have a second round effect due to micro-macro feedbacks: for example, a generous income support scheme can have effects on labor market and saving behaviors. On the other hand, macro-oriented policies or exogenous shocks have a redistributive impact that needs to be assessed if the potential implications for the political economy of the reforms is to be understood (Bourguignon and Bussolo, 2013).

As in the case of a policy change, the microsimulation approach can offer insights in two ways. First, it can provide, in a timely fashion, an ex-ante assessment of how individuals are affected by the macroeconomic changes, either actual or hypothetical. Second, it can be used to develop a counterfactual scenario to disentangle ex-post what would have happened without a given component of the macroeconomic shock. However, in order to capture the consequences of a macroeconomic shock on income distribution, a partial equilibrium setting at the micro level can be too limited. Thus, it is necessary to capture the interactions of the tax-benefit system with population heterogeneity observed at the micro level, as well as the macro changes in the fundamentals of an economy due to policy reforms or exogenous shocks.[18]

In the last decade, the growing literature has explored different ways to link micro and macro models (often belonging to the family of computational general equilibrium (CGE) models), yet the construction of a comprehensive, policy-oriented micro–macro economic model still faces many challenging issues. Although it is now quite common to

[18] This is less relevant in developing countries where the effects of social expenditures are more relevant in shaping the income distribution than tax-benefit systems, but it cannot be ignored in mature welfare states and in emerging countries due to the increasing use of conditional cash transfer programs.

see disaggregated information from micro-level data used in a macro model (i.e., using the parameters of behavioral models or the effective tax rates simulated by microsimulation models in CGE models), it is rarer to see a fully developed microsimulation model being integrated with a macro model. See Bourguignon and Bussolo (2013) for an excellent review of the different approaches.

The simplest and widely implemented way to combine micro and macro models is the top-down approach. Robilliard et al. (2008) provide an example of sequentially combining a microsimulation model with a standard multisector CGE model, not only focusing on the labor markets but also on the expenditure side taking into account the heterogeneity of consumption behavior of individuals. First, the macro model predicts the linkage variables, such as new vectors of prices, wages, and aggregate employment variables that are the consequences of a macroeconomic shock or a new policy. Second, the microsimulation model generates new individual earnings and employment status variables consistent with the aggregates from the macro model and hence simulates a new income distribution. In such a top-down approach, the potential macro feedback effects of the new situation faced by the individual are not taken into account specifically, but only through the representative households embedded in the macro model. Because it depends on the aggregation of behavior at the individual level, this approach can only provide the first-round effects of the exogenous (policy) change. Bourguignon et al. (2005) extend the top-down approach by including in the microsimulation model the behavioral reactions of individuals to the price changes predicted by the macro model.

In contrast, in the bottom-up approach, the individual behavioral changes due to a policy reform are simulated at the micro level and then aggregated to feed into the macro model as an exogenous variation in order to analyze the overall effect on the economy. Any feedback effect from the macro model back into the microeconomic behavior is ignored in this setting (Brown et al., 2009).

A more complete recursive approach is given by the combination of the two approaches through a series of iterations until effectively no further adjustments are observed, in order to take into account the feedbacks that would otherwise be ignored and to arrive at a fully integrated macro–micro model. In the macro part of such a model, the household sector is not given by a few representative households but by the micro-level sample of households representative of the whole population. Aaberge et al. (2007) is an example of the integration between a labor supply microsimulation model and a CGE model in order to assess the fiscal sustainability of the aging population in Norway. Peichl (2009) uses the same approach to evaluate a flat tax reform in Germany.

Considering the efforts needed to develop a fully integrated model, the choice of the appropriate approach depends on the research or policy question at hand, and more parsimonious models can do the job in many circumstances. Notwithstanding, a fully integrated micro–macro model, as suggested by Orcutt et al. (1976), would be an incredibly powerful tool for building counterfactuals taking into account feedback effects between the micro and macro levels and for disentangling the effect of a given macro change on individual resources and hence on income distribution.

24.3.5 Predicting Income Distribution

Using microsimulation to predict income distribution is an area of work that is fueled by the need of policymakers to have more up-to-date estimates of poverty and inequality and the effects of policy than can be supplied directly from micro-data that are usually 2–3 years out of date.

This need is particularly acute if indicators of income distribution are to be taken into account in assessments of economic and social conditions alongside aggregate economic indicators, which are generally available in a more timely way (Atkinson and Marlier, 2010; Stiglitz, 2012). Furthermore, the practice of setting targets for the future achievement of social goals is becoming more widespread. In relation to poverty and income distribution, this applies particularly in the European Union through the Europe 2020 targets for the risk of poverty and social exclusion,[19] and in the developing world through the UN Millennium Development Goal for the eradication of hunger and extreme poverty.[20] Predictions of the current situation, known as "nowcasts," are valuable indicators for measuring the direction and extent of movement toward the associated targets, along with predictions for the target date at some future point (i.e., forecasts).

The approaches for predicting income distribution also depend on the time framework of the analysis. Methods for nowcasting and forecasting are distinct in the sense that the latter must rely on assumptions or other forecasts about the economic and demographic situation, as well as the evolution of policies, rather than recent indicators, data, and known policy parameters. However, the choice of techniques is common to both, and before discussing the two time frameworks in turn, the next subsection considers a key issue: how to model changes in labor market status.

24.3.5.1 Modeling Changes in Labor Market Status

In order to capture the effects of exogenous changes in economic status on income distribution, two techniques can be implemented at the micro level. One approach is to reweight the data (Creedy, 2004; Gomulka, 1992; Merz, 1986), and another approach is to model transitions from one status to another at the individual level (Fernández Salgado et al., 2014).

Reweighting is commonly used because it is relatively straightforward to carry out and to test the effects of alternative specifications. For example, to model an increase in the unemployment rate (Immervoll et al., 2006b), the survey weights of households containing unemployed people at the time of the survey must be increased, and the

[19] http://ec.europa.eu/europe2020/targets/eu-targets.
[20] http://www.un.org/millenniumgoals/.

weights of other similar households reduced, in order to keep demographic characteristics and household structures constant in other relevant dimensions. Following this approach, Dolls et al. (2012) simulate a hypothetical unemployment shock in 19 European countries and the US in order to analyze the effectiveness of the tax and transfer systems to act as an automatic stabilizer in an economic crisis.

However, the main disadvantage of the reweighting approach, especially in the context of a rapidly changing labor market, is that it assigns the characteristics of the "old" unemployed (in the original data) to the "new" unemployed (corresponding to the current period). To the extent that the unemployed in the data were long-term unemployed, this will underestimate the number of new unemployed in receipt of unemployment insurance benefits, which are time-limited in most countries, and overestimate the extent to which incomes are lowered by unemployment. Furthermore, the unemployment shock may have affected certain industries and occupations more than others.

Another drawback of reweighting is that it can result in very high weights for some observations, which can distort the results of simulations affecting dimensions not controlled for. In addition, although the implications of alternative formulations for the empirical results are straightforward to explore, it is far less straightforward to assess the statistical properties and reliability of the weights themselves, given that for any one set of weights satisfying the calibration constraints, there are also others (Gomulka, 1992).

Moreover, reweighting does not permit the modeller to account for the interactions between changes in the individual status and different tax-benefit instruments, for example, to analyze to what extent the welfare system counterbalances income losses specifically for those who became unemployed, rather than at the aggregate level. This is possible with the second approach, which involves explicit modeling of transitions at the individual level, making use of external information about the changes occurring in a given dimension. In principle, the full range of relevant characteristics of the people affected can be taken into account. An explicit simulation allows for the detailed effects of tax and benefit policy to be captured for those making the transition. In other words, it allows the production of quasi "panel data" that tracks the same individual before and after a given change, disentangling what would have happened without the change and highlighting the interactions of the tax-benefit system with the individual sociodemographic characteristics.

Following this approach, Fernández Salgado et al. (2014) simulate individual transitions from employment to unemployment at the onset of the Great Recession in six European countries. As a consequence of the macroeconomic shock, household incomes of individuals who lose their jobs are predicted, considering the direct cushioning effect of the tax-benefit systems and the way they depend on the market income of other household members and personal/household characteristics. The comparison between incomes before and after the shock provides a way to stress-test the tax-benefit systems, assessing the relative and absolute welfare state resilience.

To date there have been few systematic comparisons of reweighting versus the explicit simulation of individual transitions. Hérault (2010) provides a comparison of results using the two methods on South African data and concludes that "the reweighting approach can constitute a good alternative when data or time constraints do not allow the use of the behavioral approach and when the production of individual level transition matrices in and out of employment is not essential" (p. 41).

24.3.5.2 Nowcasting

Tax-benefit microsimulation models have for many years been used to simulate the effects of the most recent policies so that ex ante analysis of policy reforms can take the current situation as the starting point. In doing this, it is necessary to update the input micro-data to reflect current economic and social conditions. This might be done with varying degrees of sophistication depending on the question at hand and the amount of change in relevant dimensions between the reference period of the micro-data and the reference year of the policy. Usually, information in the data on original income is updated using appropriate indexes. In addition, the sample might be reweighted to account for certain demographic and economic changes (see section 24.3.5.1). The simulated distribution of household disposable income, based on adjusted population characteristics, updated original income, and simulated taxes and benefits using current rules, is then assumed to be a reasonable representation of the current income distribution.[21]

However, in times of rapid change, two factors suggest that this approach may not be adequate. First, simple reweighting cannot generally capture major changes accurately, and income growth may vary greatly around the mean, requiring a disaggregated approach. Most obviously this applies in the case of an economic downturn and a sudden increase in unemployment with its asymmetric effects, or an upturn and an increase in employment, when, as is typically the case, the impact is uneven across the population. Secondly, it is at times of rapid change or economic crisis that policymakers particularly need to know about very recent movements in the income distribution and the current situation rather than those a few years previously. The same applies in times of growth if policymakers are concerned about some sections of the population being left behind. Furthermore, in times of crisis, fiscal stimulus or fiscal consolidation policies may play a particularly important role in reshaping income distribution. A microsimulation approach has particular advantages because it captures the specific impact of the components of these policy packages that have a direct effect on household incomes, as well as their interactions with changing market incomes. In times of rapid growth, fiscal drag will typically have distributional consequences (see Section 24.3.2.2), which will be important for policymakers to anticipate if they wish to prevent relative poverty from rising (Sutherland et al., 2008).

[21] See for example Redmond et al. (1998) and Callan et al. (1999).

Borrowing the term nowcasting from macroeconomics (see, for example, Banbura et al., 2011) the use of an extended and refined set of microsimulation methods in combination with timely macroeconomic statistics is able to provide estimates of the current income distribution using micro-data on household income which are typically 2 or more years out of date. These methods include: (a) updating market incomes from the income data year to current (or latest), using published indexes with as much disaggregation as these statistics allow and from the latest to "now" according to macro-level forecasts or assumptions; (b) simulating policy changes between the income data year to those prevailing currently; (c) adjusting data to account for important dimensions of actual labor market change between the data year and the most recently available information; (d) adjusting data to account for actual and projected demographic and other compositional changes (e.g., household composition) between the data year and "now."[22]

An early attempt to use these methods to update poverty statistics for the UK is provided by de Vos and Zaidi (1996). More recently, these methods have been used to nowcast the policy effects of the crisis in Ireland (Keane et al., 2013); to examine the distributional effects of the crisis in Greece (Matsaganis and Leventi, 2013); and to nowcast the income distribution in Ireland (Nolan et al., 2013), the UK (Brewer et al., 2013), and Italy (Brandolini et al., 2013). They have also been used for eight European Union countries to nowcast the risk of poverty, using EUROMOD (Navicke et al., 2014).[23]

A key issue for all the studies that aim to nowcast income distribution in (or on the way out of) the Great Recession, using prerecession data, is to capture labor market changes with sufficient precision. The same would apply during a period of increasing employment rates. Most of the studies cited above use reweighting to adjust for both demographic and labor market changes. The study by Navicke et al. (2014) is an exception. Holding demographic factors constant, it used explicit simulation of labor market transitions to capture the very specific and varied incidence of unemployment across the eight countries considered in the relevant period. The method is based on that employed by Figari et al. (2011c), using Labor Force Survey (LFS) statistics to establish the required number of transitions of each type according to personal characteristics. The microsimulation model, in this case EUROMOD, then selects from the available pool of people with these characteristics in the input database and changes their status accordingly. Incomes are simulated, accounting for the new status, for example, by calculating

[22] Some studies also make specific data adjustments that are relevant to the effect of policies in the projection period, such as increasing the pension age (Brewer et al., 2011).

[23] In the US, where there is also an interest in predicting current indicators, child poverty has been nowcast using a simple econometric model based on state-level current and lagged economic indicators and benefit receipt statistics (Isaacs and Healy, 2012). This is a feasible approach, rather than using microsimulation, because in the US the poverty threshold is not a function of the income distribution, as it is in the EU, and because welfare benefit receipt is in itself a good predictor of poverty, which is not the case in the EU or elsewhere in the OECD.

eligibility and entitlement to unemployment benefits for those making the transition from employment to unemployment.

24.3.5.3 Forecasting

Although nowcasts can make use of very recent indicators of economic and labor market conditions and typically project forward by only a few years, allowing slowly changing factors such as demographic composition to stay the same, forecasts project further in time and must rely on assumptions and predictions from other models. In this sense they are usually better seen as drawing out the implications for the income distribution of a particular economic/demographic scenario. For example, Marx et al. (2012) explore the implications of meeting the Europe 2020 targets for employment for indicators of risk of poverty and social exclusion, finding that the composition of any new employment is a key factor. The World Bank (2013) uses a similar approach to exploring the implications of meeting both the education and employment targets for the poverty indicators in the countries of Eastern Europe. In both cases there is no tax-benefit microsimulation, and it is assumed that tax-benefit effects are the same as those in the underlying micro-data. This is justified on the grounds that future policy reforms are difficult to predict. However, this approach neglects any interactions between sociodemographic and labor market characteristics and the tax-benefit system. Microsimulation can take account of these and, even assuming a constant tax-benefit policy structure and a constant relationship between income levels and tax-benefit parameters, would allow the automatic effects of policies on changing market incomes to be captured.

Nevertheless, as explained in Section 24.3.2.2, it is important to be aware that assumptions about the indexation of current policies can have a major effect on distributional outcomes (Sutherland et al., 2008). In some situations enough is known about the probable evolution of policies to include the discretionary tax-benefit reform effects in the predictions, as well as the automatic effects driven by changes in the circumstances of households. In the UK not only are policy reforms often announced several years in advance, but also there is detailed information available about indexation assumptions that are built into official public finance forecasts (HM Treasury, 2013; Table A1), as well as regular and detailed growth and inflation forecasts (OBR, 2013; Table 2.1) that can together be used as baseline assumptions for defining policies in the forecast year. Brewer et al. (2011) have forecast child poverty in 2020, using a combination of these types of assumptions, reweighting, and tax-benefit microsimulation. Such an approach not only provides a prediction (in this case that, given the assumptions, the UK will not meet its target for child poverty reduction in 2020) and allows the drivers of the prediction to be identified (through sensitivity analysis), but it also allows a "what would it take?" analysis to suggest what combinations of reforms and other changes would be needed to meet the target.

24.3.6 Cross Country Comparisons Using Microsimulation

Cross-country comparisons of the effects of policies naturally add value to what can be said about a single country because the broader perspective helps to provide a sense of scale and proportion. They provide the basis for assessing the robustness of results and generalizing conclusions. In addition, considering several countries within the same analysis provides a kind of "laboratory" in which to analyze the effects of similar policies in different contexts or different policies with common objectives (Sutherland, 2014). Comparisons can take several forms. At their simplest the effects of different policies or policy reforms in different countries can be analyzed side-by side. Bargain and Callan's (2010) decomposition analysis for France and Ireland is one example of this approach. Another is Avram et al. (2013) who analyze the distributional effects of fiscal consolidation measures within a given period in nine countries.

A second approach is to contrast the effects of a common, hypothetical policy reform in several countries, highlighting the relevance of the interactions of a specific policy design with population characteristics and economic conditions. Often the "reform policy" is designed to highlight features of the existing national system that it replaces or supplements. Examples include Atkinson et al. (2002) and Mantovani et al. (2007) for minimum guaranteed pensions, Levy et al. (2007a) for universal child benefits, Callan and Sutherland (1997) for basic income, Bargain and Orsini (2007) and Figari (2010) for in-work benefits, Matsaganis and Flevotomou (2008) for universal housing transfers, Figari et al. (2012b) for the taxation of imputed rent, and Paulus and Peichl (2009) for flat taxes. This type of analysis is usually complicated by the need for the reform policy to be scaled somehow if it is to have an equivalent effect in countries with different levels of income, and because of the need to consider how the reform policies should be integrated with existing national policies. Given that the starting points are different (e.g., the tax systems may treat pensions differently) the net effects will differ, too.

A third approach, which was introduced in Section 24.3.2.2, is to swap existing policies across countries in order to explore how their effects differ across different populations and economic circumstances. Examples of this kind of "policy learning" experiment include a comparison of the effectiveness of benefits for the unemployed in Belgium and the Netherlands (De Lathouwer, 1996), as well as many studies of the effectiveness of public support for children and their families: Atkinson et al. (1988) for France and the UK; Levy et al. (2007b) for Austria, Spain and the UK; Levy et al. (2009) for Poland, France, Austria, and the UK; Salanauskaite and Verbist (2013) for Lithuania, Estonia, Hungary, Slovenia, and the Czech Republic; and Popova (2014) who compares Russia with four EU countries.

Policy swap analysis can, in principle, be done using a set of national microsimulation models, side by side. But Callan and Sutherland (1997) found that the task of making models produce comparable results was formidable, even for just two (arguably) relatively similar countries (Ireland and the UK). This justified the construction of EUROMOD as a multicountry model that now covers all EU member states (see Box 24.1).

Indeed, with some exceptions, many of the studies referred to above make use of EUROMOD. As intended, this greatly facilitates cross-country comparability (particularly of the concepts used), the implementation of common reforms using common code and the mechanics of carrying out policy swaps (transferring coded policies from country A to country B). EUROMOD is designed to be as flexible as possible, allowing a huge range of assumptions to be made about cross-country equivalence of different aspects of policy simulation. One example is the treatment of non-take-up of benefits (see Section 24.4.2), and another is the default indexation of policies each year (see Section 24.3.2.2). Thus policy swapping is not a mechanical procedure. Each exercise has its own motivation and corresponding decisions to be made about which aspects of policy (and assumptions driving its impact) are to be "borrowed" from elsewhere and which are to be retained from the existing local situation.

Here, we give two empirical examples. The first is an example of side-by-side cross-country analysis using EUROMOD from Avram et al. (2013). This compares the distributional effects of the fiscal consolidation measures taken in nine European countries in the period up to 2012 from the start of the financial and economic crisis. Figure 24.6 shows the percentage change in average household income due to the measures across the (simulated) 2012 income distribution. The measures include different mixes of increases in income tax and social contributions and cuts in public pensions, other cash benefits, and public sector pay.

Four things are striking about this figure and serve to demonstrate the added value of cross-country comparisons of this type, relative to single country studies. First, the scale of the effect varies greatly across the countries (noting that the country charts are drawn to different scales, but the grid interval is uniformly 2% points), ranging from a drop in income on average from 1.6% in Italy to 11.6% in Greece. Second, the choices made by governments about which instruments to use differ across countries. Third, the incidence of the particular changes is not necessarily as one might expect a priori. For example, increases in income tax have a roughly proportional effect in many countries and are concentrated on higher-income households only in Spain and the UK, as might be expected a priori. Cuts in (contributory parental) benefits in Latvia particularly target the better off. Finally, the overall distributional effects range from broadly progressive in Greece, Spain, Latvia, and the UK to broadly regressive in Estonia.

The second example is of a policy swap, showing what would happen to child poverty in Poland under a range of child and family tax-benefit arrangements, as compared with the actual 2005 system, including a reform introduced in 2007 and the revenue-neutral alternatives offered by scaled-down versions of the Austrian, French, and UK systems of child and

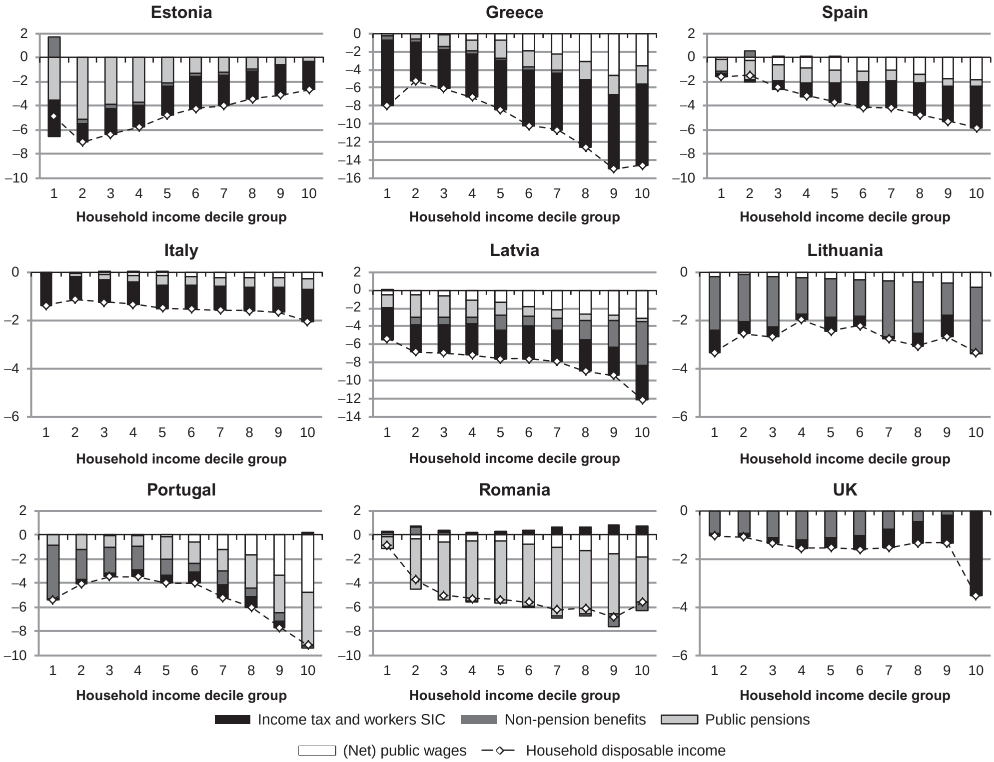


Figure 24.6 Percentage change in household disposable income due to fiscal consolidation measures 2008–2012 by household income decile group. *Notes: Deciles are based on equivalized household disposable income in 2012 in the absence of fiscal consolidation measures and are constructed using the modified OECD equivalence scale to adjust incomes for household size. The lowest income group is labeled "1" and the highest "10." The charts are drawn to different scales, but the interval between gridlines on each of them is the same. Source: Avram et al. (2013), using EUROMOD.*

family support (Levy et al., 2009). As Figure 24.7 shows, any of the alternative policy systems would have reduced child poverty by more than the actual 2007 reform (costing the same). The French and UK systems would perform especially well from this perspective.

In addition to EUROMOD, other multicountry initiatives have constructed and used microsimulation models. These include a Latin American project that built separate models using a range of software and approaches for Brazil, Chile, Guatemala, Mexico, and Uruguay (Urzúa, 2012). A WIDER project has constructed models that are available in simplified form on the web for 10 African countries.[24] To our knowledge, neither set

[24] http://african-models.wider.unu.edu/.

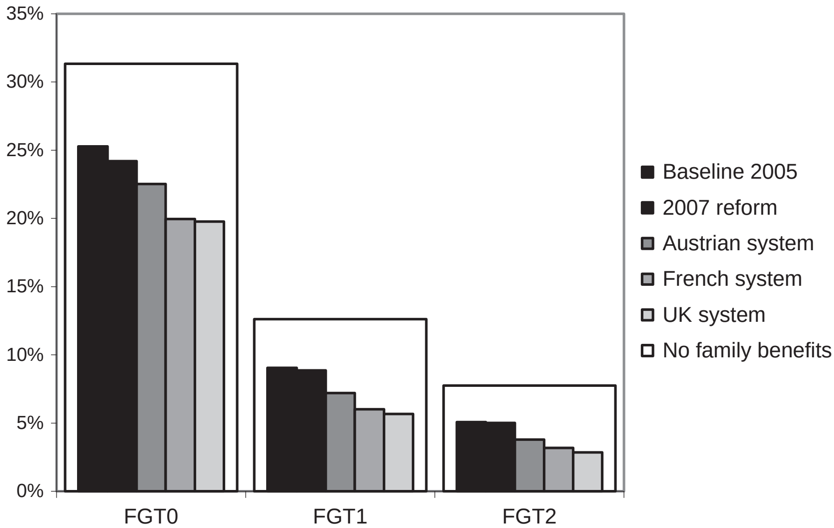


Figure 24.7 Child poverty in Poland under alternative tax-benefit strategies. *Notes: Poverty is measured using FGT indexes and 60% of median household disposable income as the poverty threshold. Source: Levy et al. (2009) using EUROMOD.*

of models has been used for cross-country comparisons of the effects of common reforms or for policy swap exercises. In contrast, there is an ongoing collaboration among some of the Balkan countries to make use of the EUROMOD platform to build models with the explicit intention of using these models for cross-country comparisons. The Serbian model SRMOD is the first completed step in this process (Randelović and Rakić, 2013), followed by the Macedonian model, MAKMOD (Blazevski et al., 2013). Similarly, the South African model, SAMOD, again using the EUROMOD platform, (Wilkinson, 2009) has been joined by a sister model for Namibia, NAMOD, with the aim, among other things, of modeling "borrowed" policies that have been successful in a South African context (Wright et al., 2014).

24.4. CHALLENGES AND LIMITATIONS

24.4.1 Reconciling Simulated Income with Recorded Income and Macro Statistics

A common problem when using micro-data from surveys for the analysis of policies and income distribution is that aggregate values (e.g., gross earnings or income taxes) do not match estimates from national accounts or other sources of macroeconomic statistics. This problem also applies to microsimulation studies based on survey data, with one exception. Tax-benefit model calculations of benefit entitlements may match administrative totals *better* than information on recorded receipt in the data, if there is a problem of underreporting of these sources of income in the survey.

Chapter 11 considers the reconciliation of household surveys and national accounts. Here, we focus on a somewhat different issue, also related to the plausibility and usability of empirical findings. This is that the simulated income distribution is not identical to the income distribution that is measured by directly using the underlying survey (or register) micro-data. Typically, measures of income inequality in microsimulated estimates, using the same micro-data and the relevant policy year, are lower. Adjustments in the simulations for the non-take-up of benefits and for tax evasion go some way to reducing the discrepancy, and these issues are discussed in Section 24.4.2. However, they appear not to be the full explanation, and it is clear that the contributory factors differ across countries. Indeed, in some countries for particular datasets and policy years, the differences are small: for example, Figari et al. (2012a) show this to be the case for four EU countries, using data from the EU Statistics on Income and Living Conditions (EU-SILC) and EUROMOD. However, this is by no means always or even often the case, and reconciling simulated and recorded estimates is an important component of both the process of building a tax-benefit model and validating the content of micro-data from surveys.

As alluded to above, there is evidence that some surveys underreport recipients of some major cash benefits, when compared with administrative statistics.[25] If the reason for this is failure to report these sources of income by recipients, then simulated benefits may perform better, generally leading to higher incomes at the bottom of the distribution and suggesting that the survey overestimates income inequality. An illustration from the UK is provided in Box 24.2.

Box 24.2 Benefit recipients in the UK: Comparing microsimulation estimates with survey responses and administrative statistics

The UK Family Resources Survey (FRS) data, which is used as input by all the main UK tax-benefit models, underestimates recipiency of some benefits by as much as 30% in the case of the pension credit, as shown in the table below.[26] Simulations using the UK component of EUROMOD based on the same data halve the scale of the discrepancy for Pension Credit. The simulations account for non-take-up by applying official estimates for each benefit and client group. Simulations are also closer to the administrative data values in the case of Housing Benefit, for which the shortfall is 2% compared with 19% in the data, but in the case of Working Tax Credit, the simulated

Continued

[25] For the UK see http://statistics.dwp.gov.uk/asd/frs/2008_09/frs_2008_09_report.pdf, Table M.6. The increasingly used practice of linking surveys to administrative sources of income data should reduce the prevalence of this problem.

[26] After adjusting the administrative statistics for recipients not living in UK private households in the case of the pension credit.

shortfall is larger. The entitlement here mainly depends on being in low paid work over the year, allowing families to meet the eligibility criteria for the working tax credit for short periods, which is not captured by the simulations based on current income and circumstances. For the other two payments shown in the table, EUROMOD over- rather than underestimates recipiency. The overestimation of Child Tax Credit recipients is to some extent explained by the administrative statistics not containing some long-term recipients of income support, whose child payments are still waiting to be migrated to the tax credit system. Most simulated and nonsimulated benefits are included in the means-test for Council Tax Benefit: its overestimation is expected to the extent that some nonsimulated benefits are underreported and tax credits are undersimulated.

Clearly, simulating receipt is not a solution in itself, and a comprehensive reconciliation needs other benefit-specific factors to be taken into account.

Numbers of recipients of selected UK benefits in the 2009–2010 tax year: estimates from Family Resources Survey (FRS), EUROMOD, and administrative statistics (thousands)

	FRS	EUROMOD	Administrative statistics	Ratio FRS/ external	Ratio EUROMOD/ external
Working Tax Credit	1,800	1,615	2,240	0.80	0.72
Child Tax Credit	3,700	4,951	4,090	0.90	1.16
Pension Credit	1,800	2,337	2,580	0.70	0.85
Housing Benefit	3,700	4,474	4,550	0.81	0.98
Council Tax Benefit	5,100	6,331	5,570	0.92	1.14

Source: EUROMOD version F6.20 with adjustments for non-take-up, using Family Resources Survey 2009/10 updated to 2010–2011 incomes.

Shortfalls in the reported receipt of means-tested welfare benefits compared with administrative information are also found in US surveys on a larger scale (Meyer et al., 2009). Wheaton (2007) uses microsimulation to calculate entitlement and then to calibrate the numbers of recipients so that they match administrative statistics. The result is a large increase in the estimated extent of poverty reduction due to the programs in question.

However, as illustrated in Box 24.2, underreporting of benefit income may not be the only source of the problem. If part of the reason for the shortfall in the survey is that benefit recipients are more likely to be nonrespondents, then microsimulation of eligibility and entitlement is unlikely to solve the problem on its own, and benefit recipiency estimates will still not match administrative information. In this case recalculation of the survey weights, including controls for characteristics that are correlated with benefit receipt and also underrepresented in the survey, may in principle provide a solution, if such characteristics can be identified and external information is available to control the process. This is not often the case.

There are many possible reasons for discrepancies in each simulated income component. Here we discuss income tax as an important example. First, survey estimates of income tax may not relate to the current year or may include only withholding taxes. Second, survey gross incomes (and hence taxes) may have been imputed from net income (see also Section 24.2.1), but their quality and consistency with calculations in the tax-benefit model are usually difficult to establish due to detailed documentation not being made available. We might also expect some discrepancies when the values are compared with fiscal data. Such comparisons need to take national specifics into account, including the nature of the tax structure and administration, as well as the questions asked in the survey. The nature of the comparison and the conclusions that are drawn also depend on whether fiscal data are available at the micro level and whether they can be matched to the survey. In addition, the fiscal data may not provide a fully reliable benchmark, especially if they are based on samples of administrative data or if the administrative process that generates them is not comprehensive or consistent. We provide a case study in Appendix B based on a published table of fiscal statistics for the UK.

Microsimulation estimates of income taxes may be over- or underestimated relative to what is shown by fiscal data. For example, income tax may be underestimated because the market incomes that make up the tax base are underreported or the survey does not adequately represent high-income taxpayers. In this case estimates of income distribution are sometimes adjusted by inflating incomes at the top of the distribution, informed by fiscal data. This is the case for the official estimates of poverty and income distribution produced by the UK Department of Work and Pensions (DWP, 2013), though the same adjustment is not (to our knowledge) applied in UK tax-benefit models. In contrast, the French model TAXIPP merges micro-data and statistics from many sources for its input database.[27] This includes information on top incomes specifically used to correctly capture the very top of the distribution and particularly the taxes paid by that section of the population (Bozio et al., 2012).

Income tax may be overestimated because of tax evasion that has not been modeled (see Section 24.4.2) or because it is not possible to model or measure the size of some tax reliefs and common avoidance measures. It may also be under- or overestimated in line with other simulated income components that are taxable. Combinations of these factors may occur, and indeed it is possible for the simulated tax aggregate to match well that from fiscal data but for the distribution of tax paid to be very different—see Appendix B for an example of this. In addition, estimates of gross income and tax liability from fiscal data may be subject to error due to tax evasion.

Time periods for income assessments are also important. In surveys that collect current income (as in the UK), which mainly use a reference time period of a month, the simulation of income tax must assume that the same monthly income was received

[27] http://www.ipp.eu/en/tools/taxipp-micro-simulation/.

all year and will not identify cases with tax liability for part of the year. However, the survey response for those with part year incomes will, at least in principle, indicate the correspondingly lower or higher tax payments, already adjusted for part-year incomes. The UK is unusual in collecting short-period current income. Most income surveys ask about annual income (in the previous year), which is the appropriate reference time period for the calculation of tax liabilities. However, it must also be used to simulate the income assessment of social assistance and other means-tested benefits for which the relevant period is generally much shorter than 1 year. This leads to fewer households being simulated to receive these benefits than shown in the data.

Generally, simulations are only as good as the underlying micro-data and, in the cases where they are necessary, as good as the imputations and adjustments that must be carried out in the absence of all the necessary information. This in turn depends on the specifics of the national benefit and tax systems as well as the quality of the data. In some circumstances it might be appropriate to calibrate and reweight to try and adjust the baseline simulated distribution of income and its components to match that given by the data directly. Generally, however, such an approach will distort the estimates of *change* due to a policy reform. A better approach is to try and understand the source of each problem and to make adjustments that can be applied in a consistent way, and with transparent assumptions, across policy scenarios. This highlights the importance not only of validation and adjustment but also of documenting the process so that users of the models and readers of model applications can make their own assessment, based on the research questions at hand.

24.4.2 Modeling Non-Take-Up and Noncompliance

One particular challenge arises with benefit non-take-up and tax noncompliance.[28] There is no natural data source with explicit information about these phenomena, and modeling each is highly context-specific. Accounting for take-up and noncompliance behavior in tax-benefit models is important because it affects estimates of fiscal aggregates (i.e., total benefit expenditures and tax revenues), but even more importantly, it can affect various parts of the income distribution in a different way. Furthermore, take-up and compliance behavior are likely to be affected by tax-benefit policy reforms and, hence, are themselves endogenous factors in the analysis. Even if microsimulation models commonly assume full take-up and compliance, this has an important implication for cross-national comparisons as results are unlikely to be consistent, as long as the prevalence and patterns of non-take-up and noncompliance vary across countries.

Benefit non-take-up refers to the situation in which those eligible for a given benefit do not successfully claim it for various reasons. This could simply be due to people not being aware of their entitlement (or even the existence of a particular form of public

[28] Benefit take-up is also referred to as welfare participation, especially in studies on the US.

support), being put off by a complex or time-consuming claiming process, or related to social stigma, such as not wanting to appear vulnerable and dependent on others' support. In an economic context, these factors can be summarized as implied costs related to take-up (Hernandez et al., 2007). Another likely key determinant is the size of the entitlement (Blundell et al., 1988), both in absolute terms and relative to other income sources and wealth of the claimant. Benefit take-up tends to be higher for universal benefits because the claiming process is simpler and the associated social stigma lower. Arguably, people are most likely to claim contributory benefits (e.g., for old age and maternity) because these are directly linked to their own previous contributions and, hence, entitlement is perceived to be more justified, while take-up of means-tested benefits tends to be lower. Therefore, assuming full take-up can distort comparisons between various benefits and make some benefits seem more effective than in fact they are. It also matters how extensive and long-established the benefit scheme is, because the benefit's scale and longevity contribute to the spread of knowledge among the population. A related phenomenon is benefit leakage, meaning that a benefit is received by those who should not be eligible. This could either indicate an unintentional error on behalf of the benefit administrator or claimant, or benefit fraud.

Studies estimating the scale and determinants of benefit take-up require information on eligibility for a given benefit and actual benefit awards. Because benefit eligibility is not directly observed (for a wider population), it must be inferred from relevant individual and household characteristics on the basis of benefit rules, and as such, it constitutes a microsimulation exercise in itself. Depending on the nature of the rules, especially when means-testing is involved, there can be complex interactions with other tax-benefit instruments, as well as with tax compliance. It is difficult to overemphasize the importance of data quality in this context, and most precise estimates can presumably be obtained with administrative data providing information as close as possible to that used by the welfare agencies, as well as actual benefit receipt (e.g., Bargain et al., 2012). For this to cover all potentially eligible people and not just claimants, it implies that agencies rely (mainly) on information from existing registries (e.g., tax records) rather than data collection from the claimants. Even then, there can still be some scope for simulation error if the claiming process involves factors such as discretion on behalf of officials awarding benefits. For example, in some countries, local social welfare offices are given a considerable level of discretion in deciding who is in greater need and, hence, more qualified for public support. On the other hand, there could be also errors made by the program administrators in the assessment of the eligibility, resulting in incorrect approval or rejection of the claim.

This type of administrative data, if it exists, is usually not accessible, and most empirical studies have relied on survey data instead. There are, however, additional challenges with survey data due to potential measurement error in the observed benefit receipts and other characteristics affecting the eligibility and the entitlement calculation (see

Section 24.4.1). For example, survey respondents may have simply forgotten the receipt of a particular benefit, associated it with an incorrect period or benefit type, or intentionally left it unreported (e.g., because of social stigma). Often, there is also a time delay between becoming entitled and receiving a first payment. Therefore, a careful assessment and cleaning of benefit data are usually required (e.g., Hancock and Barker, 2005; Matsaganis et al., 2010). Similarly, individual and household characteristics relevant for determining benefit eligibility and entitlement might be reported with error, especially other income sources and/or assets in the case of means-tested benefits. There have been only a few attempts to model the various errors explicitly (Duclos, 1995, 1997; Hernandez and Pudney, 2007; Zantomio et al., 2010).

The modeling of benefit take-up becomes even more complicated when considering the receipt of multiple benefits (e.g., Dorsett and Heady, 1991; Hancock et al., 2004), interactions with labor supply (e.g., Bingley and Walker, 1997, 2001; Keane and Moffitt, 1998; Moffitt, 1983) or dynamics in take-up behavior (e.g., Anderson and Meyer, 1997; Blank and Ruggles, 1996). Analyses combining several of these aspects are rare (e.g., Chan, 2013), and avoiding behavioral responses in other dimensions, such as labor supply, is one reason why many of the recent advances in take-up modeling have concentrated on take-up among the retired or others unable to work (e.g., Hernandez and Pudney, 2007; Pudney et al., 2006; Zantomio et al., 2010). Much of the applied research has been done for the UK and US (see above), but, among others, there are also studies for Canada (Whelan, 2010), Finland (Bargain et al., 2012), Germany (Bruckmeier and Wiemers, 2012; Riphahn, 2001), Greece and Spain (Matsaganis et al., 2010).[29] For recent reviews, see Hernanz et al. (2004) and Currie (2004).

Despite general progress with modeling take-up, it remains a challenge to deal with in microsimulation models due to the data requirements and complexities involved. Ideally, tax-benefit models should treat take-up endogenously in simulations, because policy reforms can change take-up behavior (e.g., Zantomio et al., 2010). Such attempts remain scarce (see Pudney et al., 2006). A second best approach is to predict the probability of take-up conditional on personal characteristics that are not affected by policy changes and hence remain constant in policy simulations. To predict take-up on the basis of previously estimated statistical models, the same explanatory variables need to be present in the data used for the tax-benefit model. Furthermore, take-up is highly circumstantial, and a prediction model developed for one benefit in one country is unlikely to perform satisfactorily for other benefits or countries. A simpler approach commonly used to account for incomplete benefit take-up in tax-benefit models is to assign take-up randomly among the group of eligible units for a given benefit such that the aggregate

[29] A number of US studies have focused on noncash programs such as Food Stamps (e.g. Daponte et al., 1999; Haider et al., 2003) or The Special Supplemental Nutrition Program for Women, Infants, and Children (e.g., Bitler et al., 2003).

take-up rate matches that in official statistics or previous studies (e.g., Hancock and Pudney, 2014; Redmond et al., 1998; Sutherland et al., 2008). This is obviously a rather crude approach because some people are more likely to claim than others, and, hence, it may not be sufficient to align aggregate benefit expenditure with official statistics, particularly if take-up is correlated with the level of entitlement. Another option is to link benefit entitlement to the observed receipt, which, however, seriously limits the scope for simulations.

Tax noncompliance (or tax evasion) is the other side of the coin and refers to intentional effort to lower tax liability in unlawful ways. In the context of tax-benefit models, this primarily concerns income tax and payroll tax evasion, in the form of underreporting taxable income or overreporting (income tax) deductions. Compared to benefit non take-up, this is an even more challenging issue for several reasons. First, take-up is binary by nature (i.e., an eligible person either claims or not), but tax compliance is often partial. Second, there is no single data source that would allow the precise measurement of tax evasion. Although tax records contain income reported to the tax authority, "true" income remains unobserved. Third, evading taxes may also affect how related incomes are reported to surveys. These constraints point towards the need to combine and utilize multiple data sources to study tax evasion and help to explain why hard empirical evidence at the individual level is very scarce.

Studies estimating the extent and determinants of tax noncompliance by individuals have mainly relied on audited tax records (e.g., Clotfelter, 1983; Erard, 1993, 1997; Erard and Ho, 2001; Feinstein, 1991; Martinez-Vazquez and Rider, 2005). Although tax audits are designed to detect tax noncompliance, these are not often carried out randomly and target those more likely to evade on the basis of initial screening. Repeated and extensive random tax audits, from which insights into tax evasion can be inferred for a broader population, have been primarily carried out in the US. However, even audits are unable to detect all noncompliance, especially income underreporting in which cash transactions are involved, and usually have very limited information on individual characteristics.

Surveys offer a much richer set of information on individuals but usually lack a good measure of noncompliance. Some surveys include explicit questions on compliance (e.g., Forest and Sheffrin, 2002), but given its sensitivity, the reliability of such self-reported data is unclear (Elffers et al., 1992). On the other hand, studies such as Pissarides and Weber (1989), Lyssiotou et al. (2004), and Hurst et al. (2014) have relied on indirect methods, employing econometric models that contrast surveyed income and consumption. These, however, are inevitably cruder and allow for a less detailed analysis of compliance.

Finally, laboratory experiments are common in tax compliance research, (Alm et al., 1992, 2009, 2012; Laury and Wallace, 2005). Although experiments allow one potential determinant to be isolated from the rest and for clearer conclusions to be drawn about causality, it is unclear how well conditions in the laboratory reflect actual behavior, not least as the subjects are typically students without substantial experience paying taxes.

Overall, there is substantial evidence on factors influencing people's decision to evade taxes. There are also studies showing that tax noncompliance is more prevalent for income sources that are less easily tracked by the tax authority (see Klepper and Nagin, 1989; Kleven et al., 2011). For example, the extent of underreporting income from self-employment is notably higher compared to wages and salaries because the latter are usually subject to third-party reporting (i.e., by employers), which reduces opportunities for evasion (though it does not necessarily eliminate these). Fewer studies have focused on the distributional implications of tax noncompliance (e.g., Doerrenberg and Duncan, 2013; Johns and Slemrod, 2010), some in combination with microsimulation modeling (Benedek and Lelkes, 2011; Leventi et al., 2013). For reviews of theoretical and empirical literature on tax evasion, see Andreoni et al. (1998), Slemrod (2007) and Alm (2012).

However, given the highly specific datasets that are often involved in the study of tax compliance, it is not straightforward to utilize previous findings in tax-benefit models, nor is it easy to provide one's own estimates with the type of data commonly used for microsimulation. This helps to explain why attempts to account for tax noncompliance in tax-benefit models seem to remain very limited (e.g., Ceriani et al., 2013; Matsaganis and Leventi, 2013). On the other hand, this may also reflect the fact that microsimulation studies lack details on such adjustments. Therefore, the first step towards improving the modeling of tax noncompliance (as well as benefit take-up) is increasing transparency about how this is handled (if at all) in existing models and studies.

24.4.3 Assessing the Reliability of Microsimulation Estimates

The overall credibility of a microsimulation model in simulating the effects of a given tax-benefit policy encompasses different aspects, some of which are interrelated, and include the application of "sound principles of inference in the estimation, testing and validation" (Klevmarken, 2002).

First, the reliability of a microsimulation model is closely tied to its validation and transparency, which are indicated by the extent to which solid documentation exists for the internal features of the model and the validation of the results against external statistics. Unfortunately, a high level of transparency does not characterize many of the microsimulation models used in the academic and policy literature, which tend to be "black boxes." Good practice is to provide a detailed description of all tax-benefit components simulated, including details of assumptions used, as well as information about the input data and related transformations or imputations. Documented validation of the output against external statistics on benefit recipients and taxpayers and total expenditure/revenue is also an important component of the informed use of microsimulation models.

Nevertheless, such validation is not a comprehensive assessment for three reasons. First, as illustrated in Section 24.4.1, microsimulation estimates and the information available in official statistics may not be comparable conceptually. Second, in some countries, limited external information is available, and in all it is rarely available without a time delay. Third, although it is possible to validate results for existing and past systems, it is usually not possible to find independent estimates of the effects of policy reforms. A correct baseline does not ensure that the model or its input data can correctly estimate the effect of a reform.

In addition, as mentioned by Wolf (2004), a persistent failure of most microsimulation applications is the lack of recognition of the degree of statistical uncertainty associated with the results, some of which is inherent in the sampling process that underlies the input micro-data and some of which is propagated from simulation errors and estimated parameters. The accuracy of the underlying data, the correct and detailed representation of the tax-benefit rules, and the actual implementation of the policy parameters in the simulation code determine the point-estimate of the simulated policy. Nevertheless, the correct interpretation of the results should take into account their statistical inference—an aspect often neglected in the microsimulation literature—which also depends on the nature of the model and whether it is purely deterministic or also involves probabilistic or econometric specifications.

To start with, simulations are subject to the same degree of sampling error, measurement error, and misreporting as any other analysis based on survey data. On the one hand, as discussed in Section 24.4.1, simulations can improve the accuracy of results by simulating the exact rules rather than relying on observed values that might be misreported. On the other hand, the simulation process can introduce other sources of errors due to, for example, approximations in the simulation of tax benefit rules, adjustments for noncompliance or non-take-up, updating of monetary parameters and sociodemographic characteristics to the simulation year, or ignoring behavioral responses or market adjustments.

In the case of simulation of the first-order effects of policy changes, Goedemé et al. (2013) argue that the lack of attention to the statistical significance of the results is undesirable and unjustified due to the availability of standard routines embedded in most standard statistical software. Moreover, when comparing the statistics related to different scenarios, they show the importance of taking into account not only the sampling variance of the separate point estimates but also the covariance between simulated and baseline statistics which are based on the same underlying sample. This can lead to a generally high degree of precision for estimates of the effects of a reform on a particular statistic of interest.

The situation is much less straightforward in the case of more complex simulations involving revenue-neutral reforms or behavioral reactions that add additional sources of uncertainty due to the use of estimated wage rates for constructing the budget set and the preference parameters estimated using econometric models. Despite the growing literature on estimating the labor supply effects of policy changes (see Section 24.3.3), there are only a few examples of studies focusing on the analytical properties of the sampling distribution of the microsimulation outcomes that are affected by simulation uncertainty and estimation uncertainty. The former stems from the simulated choice set that can be different from the one that an agent would choose in reality. The estimation variability comes from the sampling variability of the estimated parameters of the labor supply model (Aaberge et al., 2000). Pudney and Sutherland (1994) derived the asymptotic sampling properties of the most important statistics usually reported in microsimulation studies, taking into account the additional uncertainty introduced by the imposition of revenue neutrality in the construction of the confidence intervals. Pudney and Sutherland (1996) augmented the previous analysis, deriving analytically the asymptotically valid confidence intervals of a number of statistics, allowing for errors associated with sampling variability, econometric estimation of parameters of a multinomial logit model of female labor supply, and stochastic simulation in the calculations. They concluded that sampling error is the largest source of uncertainty, but parameter estimation errors may add additional uncertainty that undermines the practical use of such behavioral models.

The complexity of the analytical solution associated with very detailed microsimulation models, rather complex policy simulation and sophisticated econometric models, has lead to the use of more tractable empirical approaches. Creedy et al. (2007) opted for a simulation approach to approximate the sampling distribution of statistics of interest based on the sampling distribution of the estimated parameters. The approach relies on a number of draws from the parameter distribution of the underlying behavioral model. Moreover, they suggest a simpler and more practical approach in which the functional form of the sampling distribution is assumed to be normal, requiring a small number of draws from the parameter distribution and leading to generally accurate results.

Furthermore, to avoid having to assume the normal distribution for stochastic terms, and exploiting the increasingly available computer power, assessing the statistical reliability of the estimates now commonly relies on resampling methods such as the bootstrap, which allows one to obtain a set of replicated econometric estimates used in one or more simulation runs. The variance of the replicated estimates is then used to capture the variability of the statistics of interest. Although the additional uncertainty added by behavioral modeling is not found to be critical for most analysis (e.g., Bargain et al., 2014), there are reasons for concern when the estimates refer to specific small demographic groups, and further developments in this research area are needed.

24.5. BROADENING THE SCOPE

24.5.1 Extended Income, Consumption and Indirect Taxes

Although disposable income is the most-used indicator of living standard, it is widely recognized that economic well-being is a multidimensional concept (see Chapter 2).

The economic value of the consumption of goods and services, including interhousehold transfers, in-kind benefits, and homeowners' imputed rent related to the main accommodation, is often considered a better indicator than income when measuring individual well-being on both theoretical and pragmatic grounds (Meyer and Sullivan, 2011). The exclusion of consumption expenditure and noncash income from empirical studies of the redistributive effect of tax-benefit systems might also hamper cross-country comparison given the different degree of monetization of the economy across countries. Moreover, the distributional impact of policy changes may be rather different if noncash incomes and indirect taxes are included, with important implications for the design of policies aiming to fight poverty and social exclusion, because such an omission may lead to imperfect targeting and misallocation of resources. Notwithstanding their importance, most microsimulation models do not include either in-kind benefits or indirect taxes, mainly due to data limitations.

In European countries, in-kind benefits, such as services related to child and elderly care, education, health, and public housing, represent about half of welfare state support and contribute to reducing the inequality otherwise observed in the cash income distribution. The economic value of public in-kind benefits can be imputed at individual and household levels on the basis of per capita spending, considering the average cost of public services (such as providing care and education services), the gain from paying below-market rent or no rent at all for public housing, or the risk-related insurance value approach that considers public health care services equivalent to purchasing an insurance policy with the same cost for individuals who have the same sociodemographic characteristics. See Aaberge et al. (2010b) and Paulus et al. (2010) for empirical evidence across European countries and for methodological insights on the derivation of needs-adjusted equivalence scales that are more appropriate for extended income. However, survey data usually do not include enough information to simulate changes in the value of the benefit due to policy reforms, nor do they take into account the real utilization by the individual, the quality of the public service, or the discretion in the provision usually applied by local authorities (Aaberge et al., 2010a).

A more comprehensive measure of individual command over resources should include the income value of home ownership as well. This is because the consumption opportunities of homeowners (or individuals living in reduced or free rent housing) differ from those of other individuals due to the imputed rent that represents what they would pay if they lived in accommodation rented at market prices. The inclusion of imputed rent in microsimulation models is becoming more common due to the refinement of different methods for deriving a measure of imputed rent (Frick et al., 2010) and also a renewed interest in property taxation. From a cross-country perspective, Figari et al. (2012b) analyze the extent to which including imputed rent in taxable income affects the short-run distribution of income and work incentives, showing a small inequality-reducing effect together with a nontrivial increase in tax revenue. This offers the opportunity to shift the fiscal burden away from labor and to increase the incentive for low-income individuals to work.

Indirect taxes typically represent around 30% of government revenue. With only a few exceptions, household income surveys providing input data for microsimulation models do not include detailed information on expenditures either, preventing micro-level analysis of the combined effect of direct and indirect taxation. The solution usually adopted to overcome this data limitation is to impute information on expenditures into income surveys (Sutherland et al., 2002). Decoster et al. (2010, 2011) provide a thoughtful discussion of the methodological challenges and a detailed explanation of the procedure implemented in the context of EUROMOD for a number of European countries. Detailed information on expenditure at the household level is derived from national expenditure surveys, with goods usually aggregated according to the Classification of Individual Consumption by Purpose (COICOP), identifying, for example, aggregates such as food, private transport, and durables. The value of each aggregate of expenditure is imputed into income surveys by means of parametric Engel curves based on disposable income and a set of common socioeconomic characteristics present both in income and expenditure datasets. In order to prevent an unsatisfactory matching quality in the tails of the income-expenditure distributions, a two-step matching procedure can be implemented by first estimating the total expenditures and total durable expenditures upon disposable income and sociodemographic characteristics and then predicting the budget share of each COICOP category of goods. Moreover, the matching procedure takes into account the individual propensity for some activities, such as smoking, renting, using public transportation, and education services, which are not consumed by a large majority of individuals. Individual indirect tax liability is then simulated according to the legislation in place in each country, considering a weighted average tax rate for each COICOP category of goods imputed in the data.

Most microsimulation models that include the simulation of indirect taxes rely on the assumption of fixed producer prices, with indirect taxes fully passed to the final price paid by the consumer. To relax such an assumption one should go beyond a partial equilibrium framework and link the microsimulation models to macro models (see Section 24.3.4) in order to consider the producer and consumer responses to specific reforms or economy-wide shocks. There is some variety in the ways in which the models deal with the estimation of changes in spending patterns due to the simulated reforms (Capéau et al., 2014). Some models simulate only a nonbehavioral first round impact (i.e., quantities or expenditures are kept fixed at the initial level), and others estimate partial behavioral reactions taking into account the income effect on demand for goods and services by means of Engel curves (Decoster et al., 2010) or even full demand systems accounting for the real income effect and the relative price effects (Abramovsky et al., 2012).

The inclusion of indirect taxes also raises the question of how to measure their incidence. Table 24.1 shows the incidence of indirect tax payments for three European

Table 24.1 Incidence of indirect tax payments

	As % of disposable income			As % of expenditures		
Income decile	**Belgium**	**Greece**	**UK**	**Belgium**	**Greece**	**UK**
1	15.3	37.7	20.2	11.3	13.5	13.9
2	12.0	23.4	13.5	11.8	13.9	14.0
3	11.7	19.8	12.6	12.1	14.3	13.8
4	11.6	18.4	12.4	12.5	14.2	13.8
5	11.4	17.6	11.8	12.7	14.2	14.1
6	11.0	16.0	11.6	12.8	14.1	14.3
7	10.9	16.0	11.1	13.1	14.6	14.5
8	10.8	14.9	10.7	13.3	14.2	14.7
9	10.5	14.2	9.9	13.5	14.3	14.6
10	9.9	11.9	8.2	13.9	14.1	14.4
Total	11.1	16.0	10.8	12.9	14.2	14.3

Notes: Decile groups are formed by ranking individuals according to equivalized household disposable income, using the modified OECD equivalence scale.
Source: Figari and Paulus (2013), based on EUROMOD.

countries expressed as a percentage of disposable income and as a percentage of expenditure, by decile of equivalized disposable income. In the first case (see the left panel of Table 24.1), the regressivity of indirect tax payments is clear: poorer individuals pay a larger proportion of their income in indirect taxes compared to richer individuals, mainly due to a larger propensity to consume or even dissaving reflected by average expenditures exceeding incomes for the individuals at the bottom of the income distribution (Decoster et al., 2010). However, survey data might suffer from measurement error, in particular from income underrecording (Brewer and O'Dea, 2012), which could give a misleading snapshot of the income-consumption pattern at the bottom of the income distribution. In the second case (i.e., the right panel of Table 24.1), indirect tax payments are progressive, and poorer individuals pay a slightly smaller proportion of their total expenditure in VAT and excises compared to richer individuals. The main reason for this is that the goods that are exempt from VAT or subject to a lower rate (e.g., food, energy, domestic fuel, children's clothing) represent a much larger share of the total spending of poorer individuals than of richer individuals (Figari and Paulus, 2013). The distributional pattern of the indirect taxes being regressive with respect to disposable income and proportional or progressive with respect to expenditure reinforces, on empirical grounds, the importance of the choice of the measurement stick that should be used as a benchmark in the welfare analysis (Capéau et al., 2014; Decoster et al., 2010).

The potential of microsimulation models that are capable of simulating direct and indirect taxes within the same framework is reinforced by the renewed interest in the tax shift from direct to indirect taxation in order to enhance the efficiency of the tax system (Decoster and Van Camp, 2001; Decoster et al., 2010). In particular, microsimulation models have been used to assess the distributional consequences of a "fiscal devaluation," a revenue-neutral shift from payroll taxes toward value-added taxes that might induce a reduction in labor costs, an increase in net exports, and a compression of imports, with an overall improvement in the trade balance (de Mooij and Keen, 2013; European Commission, 2013).

Two general considerations arise from the use of microsimulation models for the analysis of the redistributive effects of indirect taxes. On the one hand, the actual degree of regressivity of indirect taxes might be less than that observed if surveys tend to underreport income more than consumption at the bottom of the income distribution (Brewer and O'Dea, 2012; Meyer and Sullivan, 2011). On the other hand, a more systematic use of simulated income values, as generated by a microsimulation model rather than as observed in the data, can help in solving the underreporting of income values, closing the gap between reported income and consumption and providing a more robust indicator of living standards for those with a low level of resources.

24.5.2 Dynamic Microsimulation and Lifetime Redistribution

The importance of investigating the "long-range character" of public policies was already highlighted by Guy Orcutt in the 1950s (Orcutt, 1957) and pioneered through his work in the 1970s on DYNASIM, a dynamic microsimulation model of the US designed to analyze the long-term consequences of retirement and aging issues (Orcutt et al., 1976). A number of reviews survey the existing dynamic microsimulation models, the methodological challenges, and the types of uses, providing an overall picture of the evolution of the state of play and future research directions for interested readers (Gupta and Harding, 2007; Harding, 1993, 1996b; Harding and Gupta, 2007; Li and O'Donoghoue, 2013; Mitton et al., 2000).

Dynamic microsimulation models extend the time frame of the analysis in order to address the long-term distributional consequences of policy changes, widening the perspective of the effects of the policies to encompass the individual lifetime and addressing questions about intrapersonal redistribution over the lifecycle (Harding, 1993). Dynamic microsimulation models typically aim to capture two main factors that shape the income distribution in a long-term perspective. First, they cover the changing structure of the population due to evolving individual and household characteristics (e.g., age, education, household composition) and life events (e.g., marriage, household formation, birth, migration). Second, they capture the interaction of market mechanisms (e.g., labor market participation, earnings levels) and the tax-benefit system with such characteristics in each point in time.

In particular, they are useful tools to analyze: (i) the performance of long-term policies such as pensions and other social insurance programs such as health and long-term care, (ii) the consequences of different demographic scenarios, (iii) the evolution of intertemporal processes and behaviors such as wealth accumulation and intergenerational transfers, and (iv) the geographical trend of social and economic activities if dynamic microsimulation models are supplemented with spatial information (Bourguignon and Bussolo, 2013; Li and O'Donoghoue, 2013).

The methodological challenges behind a microsimulation model depend on the scope of the events taken into account and the methodology used to age the population of interest through the period of analysis. The aging process can be either static or dynamic. With the static aging method, the individual observations are reweighted to match existing or hypothetical projections of variables of interest. The approach is relatively straightforward, but it can become unsatisfactory if the number of variables to be considered simultaneously is large or if one is interested in following individual transitions from one point in time to the next (see also Section 24.3.5.1). The dynamic aging method builds up a synthetic longitudinal dataset by simulating individual transition probabilities conditioned on past history and cohort constraints that take into account the evolution of the sociodemographic characteristics of interest through the time horizon of the analysis (Klevmarken, 2008). The major source of information for the estimation of the dynamic processes is derived from longitudinal data available in most developed countries, although often the duration of the panel is not long enough to observe transitions for large samples of individuals, the main exceptions being the long panel data available in Australia, Germany, the UK, and the US. Transitions can be estimated through reduced form models that incorporate deterministic and stochastic components, or they can be simulated, taking into account behavioral reactions of individuals to other changes that occurred at the same time, based on individual preferences estimated through structural models that take into account the endogeneity of some individual transition probabilities (see Section 24.3.3).

The aging of individual and household characteristics can be implemented as a discrete or continuous process. The former is usually built around yearly time intervals; it is more straightforward but implies that some simulated events might not respect the real sequence. The latter is based on survival functions that consider the joint hazard of occurrence of the simulated events.

In principle, dynamic microsimulation models allow for analysis that is more in line with the theoretical arguments in favor of a lifetime approach to the analysis of the redistributive effects of tax-benefit systems, as developed in the welfare economics literature (Creedy, 1999a). Nelissen (1998) is one of the few examples where the annual and lifetime redistributive effects of the social security system (here for the Netherlands) are analyzed simultaneously, making use of the same microsimulation model that guarantees comparable simulations of the tax-benefit system in place over a long period of time. In line with other research (e.g., Harding, 1993), Nelissen (1998) finds that the lifetime redistributive effect is considerably smaller than the annual incidence, with important policy implications due to the different incidence of various pension schemes on different generations.

Due to the complexity of the aging process, early dynamic microsimulation models tended not to address the long term implications of policy and policy change on income distribution as a whole (i.e., population-based models) but rather focused on specific cohorts of the population (cohort models). Nowadays such a distinction is less significant due to the improvements in the modeling set up as well as major improvements in available computing power. However, despite the improvements in dynamic microsimulation modeling, such models are often perceived as black boxes, making it difficult to understand and appreciate their properties. In particular, the lack of good economic theory and sound econometric inference methods are thought to contribute to a sceptical view of these models by the economics profession (Klevmarken, 2008).

Two particular research developments characterize the dynamic microsimulation field. First, this is an area where international collaborations are emerging in an attempt to reduce the efforts needed to build very complex models. The Life-Cycle Income Analysis Model (LIAM) stands out as a viable option to provide a general framework for the construction of new dynamic microsimulation models (O'Donoghue et al., 2009) and to be linked to EUROMOD (and other modular-based microsimulation models) in order to exploit the existing parameterization of tax-benefit systems for the European countries (Liégeois and Dekkers, 2014). Second, most dynamic microsimulation models do not include macro feedback effects and do not have market clearing mechanisms that would require ambitious links to macro models (Bourguignon and Bussolo, 2013). However, due to the number and complexity of the interactions between many social and economic variables involved in the modeling, the integration between dynamic micro and macro models could introduce too much uncertainty in the results to make them useful in a policy context (Li and O'Donoghoue, 2013).

24.5.3 Crossing Boundaries: Subnational and Supranational Modeling

The natural territorial scope for a microsimulation model is a country or nation. This is because in most countries some or all of the tax-benefit system is legislated and administered nationally; the micro-data used as an input dataset are representative at the national level; the other data used to update, adjust, and validate the model are usually made available at national level; and the economy and society are usually assumed to exist and operate at this level. However, in some countries, policies can vary across regions, sometimes following from (or accompanied by) major differences in politics, history, and economic and social characteristics. In some cases, the data that are especially suitable as the basis for microsimulation modeling are only available for one region. For these reasons, models may exist for single regions, or national models may be able to capture regional differences in policy. Examples of regional or subnational models include Decancq et al. (2012) for Flanders (Belgium) and Azzolini et al. (2014) for Trentino (Italy); both are based on the EUROMOD framework, and the latter exploits a rich dataset that combines administrative and survey data. Examples of national modeling exercises that capture extensive regional differences in policies include Cantó et al. (2014) for Spain.

If the micro-data are representative of each region, then the national model can operate as a federation of regional models, also capturing any national policy competencies. As well as simulating the appropriate policy rules regardless of location (many models for countries with regional policy variation simply opt to simulate policies from a single "representative" region), these federal models can identify the implied flows of resources (redistribution) between regions as well as within them, given budget constraints at either national or regional levels. In the US, the most comprehensive in terms of policy coverage is the long-standing microsimulation model, TRIM3, which simulates welfare programs, as well as taxes and regional variation in programs, making use of a common national input dataset: the Current Population Survey (CPS) Annual Social and Economic Supplement (ASEC).[30] See, for example, Wheaton et al. (2011) who compare the effects of policies on poverty across three U.S. states. For Canada, the microsimulation model SPSD/M has been linked to a regional input-output model in order to capture some of the indirect effects of national or provincial tax-benefit policy changes at the provincial level (Cameron and Ezzeddin, 2000).

In the European Union, policies in the 28 member states vary in structure and purpose to a much greater degree than they do across US states. Although the EU-SILC data is output-harmonized by Eurostat, it is far from ideal as an input database for a microsimulation model (Figari et al., 2007), and significant amounts of nationally specific adjustments are needed to provide the input data for EUROMOD, the only EU-wide model (see Box 24.1). Indeed, although the supranational administration of the EU has no relevant policy-making powers (at the time of writing), analysis that considers the EU (or the eurozone) as a whole is highly relevant to approaching the design of tax-benefit policy measures to encourage economic stabilization and social cohesion. Analogously to regionalized national models, EUROMOD is able to draw out the implications of potential EU-level policy reforms for both between- and within-country redistribution (Levy et al., 2013), policy harmonization, and stabilization (Bargain et al., 2013a), as well as for the EU income distribution.

At the other extreme, microsimulation methods have been used to estimate income distribution and other indicators for small areas. This relies on spatial microsimulation techniques (Tanton and Edwards, 2013) or, more commonly, reweighting national or regional micro-data so that key characteristics match those from census data for the small area (Tanton et al., 2011). In the developed world, policymakers generally use these models to predict the demand for services such as care facilities (for example, Lymer et al. (2009) for Australia and Wu and Birkin (2013) for the UK). In circumstances where the census data provide a good indication of income levels, such as in Australia, they have

[30] http://trim3.urban.org.

also been used to provide small area estimates of income distribution and its components (Tanton et al., 2009). Linkage of the census with household budget survey data in the UK has been used to estimate the small area effects of an increase in VAT (Anderson et al., 2014). A similar method known as "poverty mapping" has been applied to developing countries by Elbers et al. (2003), using household budget surveys and census micro-data in order to monitor the geographic concentration of poverty and to evaluate geographic targeting of the poor as a way of rebalancing growing welfare disparities between geographic areas. For the use of the model for Vietnam, see Lanjouw et al. (2013).

24.6. CONCLUSIONS AND OUTLOOK FOR THE FUTURE

24.6.1 What Has Been Achieved So Far?

Tax-benefit modeling is now in widespread use to provide evidence in the policy-making process. Tax-benefit models are used within governments to provide costings of policy reforms and impact assessments of distributional and incentive effects. They are used to assess progress towards meeting targets within relevant policy domains (and may be used to set feasible targets in the first place). They are used to explore the implications of alternative reform options. Other participants in the policy-making process (opposition political parties, special interest groups, NGOs, international organizations, and civil society generally) may also put forward their own perspectives and alternative proposals on the basis of microsimulation analysis. All of them may draw on the growing body of microsimulation-informed economic analysis from academic research. Within academia, microsimulation is also an accepted and recognized part of the toolbox in applied public economics, other branches of applied economics, and other disciplines, such as quantitative social policy, sociology, and political science. Evidence for this is provided by the increasing frequency of publication of articles making use of microsimulation in mainstream journals, as is clear from the references included in this chapter, and reliance on microsimulation analysis in the economic debate, as illustrated by Mirrlees et al. (2010).

Microsimulation modeling provides an opportunity for fruitful links between the policy-making and academic communities. There are many instances in which methodological developments within academic policy-focused research have provided new and more sophisticated tools that can be adopted for use by policy-making institutions. One example is the modeling of labor supply responses, which is increasingly included in microsimulation models used by government agencies. There are also instances in which innovation has taken place within government agencies in response to particular policy needs, as well as instances of the analytical needs of policymakers providing the impetus for academic developments. One example from the European Union is the adoption of social targets for Europe 2020 and the need to develop methods of forecasting micro-level indicators. Forging such links can bring additional benefit in the form of more open channels of communication with the official producers and providers of micro-data about the data requirements of microsimulation models and the potential benefits for policy-making.

In our view there are four major strands of technical/methodological achievement and ongoing progress in the use of tax-benefit microsimulation for the analysis of policy and income inequality. A *formal framework* for disentangling the effect of policies on income distribution is an important step toward better understanding how various studies have approached measuring these effects and their consistency. A coherent framework can no doubt greatly increase the clarity and transparency of microsimulation studies and facilitate links with other relevant methodological literature. The devil is in the details and microsimulation modeling offers these in abundance.[31]

Behavioral microsimulation is no longer limited to the academic sphere, and it has an increasing impact on policy-motivated analysis. Further developments of behavioral models in terms of policy scope (e.g., extending economic modeling to cover areas such as housing, mobility, and saving) and their robustness based on the comparison with ex-post evaluation studies may strengthen their role in the policy and economic debate. Moreover, the cross-fertilization between the analytical and the computational approach to the optimal taxation problem based on behavioral microsimulation models could reinforce the link between public finance theory and applied research.

The analysis of tax-benefit policies with a clear impact on the labor market participation and the evaluation of the impact of macroeconomic shocks would clearly benefit from the availability of counterfactuals that consider *feedback effects between the micro and macro level*. A fully integrated micro–macro model, although daunting in terms of the time and resources required to create it, is potentially an incredibly powerful tool for moving beyond the partial equilibrium framework in which microsimulation models operate, for disentangling the effects of macro changes on individual resources, and for extending the policy scope of the analysis through the linkage to environmental models. However, the practical, conceptual, and methodological challenges are formidable. Even so, falling short of full model integration, improving methods of linking microsimulation analysis to macroeconomic data in various ways has been, and remains, an important part of the developing toolbox.

Cross-country comparisons of policy effects, and especially policy swap analysis, inform our understanding of the variation in the effects of policies in different economic and sociodemographic contexts, and, at the same time, these comparisons offer the opportunity for cross-country "policy learning." The development of EUROMOD, and other multicountry models, has facilitated this type of analysis, while maintaining comparability of concepts and measurement and consistency in the operation of policy rules. There is

[31] Furthermore, as Spielauer (2011, p. 18) has reflected: "If beauty is to be found in simplicity and mathematical elegance (a view not uncommon in mainstream economics), microsimulation models violate all rules of aesthetics."

potential to extend the approach to global regions other than the EU, such as southern Africa, Latin America, or the Balkan region (where, arguably, policy learning is most relevant). There is also potential to extend beyond the EU to include all OECD countries to aid comparisons, for example, between the EU and the US.

There is room for improvement and for development in two key areas. The first relates to the data and methods that are available for input into and adoption by microsimulation models. Our understanding of how available micro-data may be improved and reconciled with other information, as well as the potential of new forms and sources of data that may improve the quality and scope of simulation or facilitate linkage with other models (macro, environmental, etc.), are areas for attention. In terms of methodological improvements, more attention is clearly needed to assess statistical significance and reliability of results obtained with microsimulation models drawing on various statistical methods.

The second area for improvement relates to the organization of microsimulation activities. There is much duplication of effort (with many models doing the same or similar things in some countries), combined with problems of lack of transparency (i.e., lack of documentation, results that are not reproducible by others). Furthermore, most existing models are not made available or accessible to the people who might make use of them. The final two subsections explore the outlook for microsimulation and policy analysis along these two dimensions.

24.6.2 Data and Methodological Developments

Microsimulation models require access to appropriate and good quality micro-datasets that are themselves well-documented and validated against independent information. The trend toward making more use of register (administrative) data to supply information on income receipt (and in some cases many other variables) is welcome in the sense that it reduces measurement problems and underreporting and potentially frees up resources (e.g., survey interview time) for the collection of more or better quality data in other dimensions. At the same time, such linkage may introduce new problems. It may delay the delivery time of the micro-data if there are limits on the speed of obtaining and processing administrative information. Use of administrative information may also raise new concerns about data confidentiality, which may result in additional restrictions on the ways in which the datasets can be accessed and by whom. There seem to be trade-offs between using high-precision data and widespread access.

Technological developments may offer possible ways around these trade-offs, if models and their micro-level data (both input and output) are housed on a suitably secure server and accessed remotely. This is a mode of working that was pioneered for income distribution analysis by LIS[32] and, in spite of the additional complexities associated with

[32] http://www.lisdatacenter.org/.

microsimulation modeling, has also been successfully deployed in a few other cases. These include the WIDER African models, as well as two adaptations of national components of EUROMOD: Mefisto for Flanders (Decancq et al., 2012) and Soresi for Austria.[33] In each of these cases, the broad aim of the models is to provide access to modeling capacity by civil society, with the simulation and output options offered to users structured and restricted accordingly. More critically in this context, in each case the providers of the input micro-data have given permission for such access over the web. It remains to be seen whether it will be possible to make use of high-precision administrative data in this way. Even so, there would be other technical and pedagogical challenges to be overcome in offering to the public the full flexibility of a model like EUROMOD using remote access.

More generally there is potential to extend the policy scope and applicability of microsimulation models through the statistical linkage of data from different sources. Given the increasing complexity of tax-benefit systems that operate through direct and indirect taxes, wealth and property taxes, and cash and noncash benefits, microsimulation models can help in understanding the overall effect on individual material well-being only if more comprehensive surveys become available, cross-links between various administrative datasets are utilized further, or systematic and rigorous matching procedures are implemented and documented. A prime example is the analysis of the effects of indirect taxes, because any conclusion about the incidence and regressivity of taxes can be easily biased by the data inconsistency observed, in particular, at the tails of the income distribution (see Decoster et al., 2010; Brewer and O'Dea, 2012).

Finally, making progress on many of the technical challenges associated with microsimulation modeling, most notably the modeling of take-up and compliance behavior, is also inhibited by lack of suitable data. For example, nonreceipt of a benefit entitlement may be explained in many ways, ranging from (among other causes) measurement error in the survey responses, lack of information about eligibility on the part of a nonclaimant, or a decision not to claim due to the costs of claiming. It is likely that the relative importance of each factor varies with national context and specific benefit. Accurate modeling of the probability of taking up (i.e., receiving, given positive entitlement) a particular benefit, in principle, needs to take any one or many possible causes into account, which would typically be demanding in terms of the data requirements. Modeling of tax noncompliance at the individual level is even more demanding given the concealed nature of such activities and a potentially wider range of possible factors and interactions at play. Progress in these areas can therefore be expected to be patchy and uneven, depending on the specific problems and the data possibilities.

[33] http://soresi.bmask.gv.at/Mefisto/.

24.6.3 The Case for a Collaborative Approach

Few models are accessible beyond their producers. This leads to a proliferation of many similar models and the (largely wasteful) duplication of effort that this involves. It also limits access to models because building from scratch is time-consuming and requires specialist skills; there are significant barriers to entry. Furthermore, the need to provide in the public domain documentation or validation of models that are essentially private to their producers is rarely acted upon. This lack of transparency inhibits proper evaluation of microsimulation-based studies, and lack of access inhibits the reproducibility of microsimulation analyses. Together, these factors may reduce the chances of microsimulation-based studies being published in the top scientific journals. As Wolfson (2009, p. 29) says:

> *.... microsimulation modelling still has not achieved the kind of scientific status it deserves. One reason is that many potential users are concerned about the 'black box' nature of microsimulation models. An important step, therefore, is for microsimulation modelling to become a 'glass box' activity, including for example public availability of the model and open source code.*

Models are also expensive to maintain and keep up to date. If there were fewer, better models that were made generally accessible, this would improve efficiency and quality. A collaborative approach would also bring the various types of use and user closer together and, with the appropriate level of (technical) model flexibility, could also facilitate innovations such as model linkages. EUROMOD and TAXSIM provide two rather different examples of models that already take this approach. EUROMOD makes available both tax-benefit codes and input data to anyone with permission to access the original micro-data sources, while TAXSIM provides online access to the tax calculator that may be linked to input data of the user's own choosing.[34]

Of course, there are also good reasons why microsimulation models are developed as individually or institutionally private investments. In some cases the necessary micro-data cannot be made available more widely (e.g., in the case of government models, especially those using administrative data). In the academic sphere, there are few incentives to share technical developments as public goods in the matter suggested, especially if they embody a large time investment and if they do not themselves attract academic reward.

If the benefits of an open and collaborative approach are to be realized the main challenges are to find ways of organizing and funding arrangements that account for the long-term investment aspect, due to the need to maintain models, as well as engage in initial construction. This would include developing an incentive structure that recognized the academic value of the work done on the "public good" research infrastructure, while

[34] In addition, there are the web-based, simplified models referred to above, as well as other web-based developments, including the French model OpenFisca http://www.openfisca.fr/.

eliciting contributions in some form from the users of the models who might otherwise "free ride." In the end, cooperation within the microsimulation community and particularly between academic researchers and policy makers will contribute to the integration of microsimulation for policy analysis into the mainstream of economic policy-making (Atkinson, 2009).

ACKNOWLEDGMENTS

We wish to thank Tony Atkinson, François Bourguignon, and Brian Nolan for their comments and suggestions on the early drafts of this chapter. We are also grateful to Paola De Agostini, John Creedy, Mathias Dolls, Carlo Fiorio, Horacio Levy, Marcello Morciano, Andreas Peichl, Iva Tasseva, and Alberto Tumino for comments and useful discussions, as well as information or permission to make use of their analysis. Paulus and Sutherland acknowledge the support for this work from the core funding of the Research Centre on Micro-Social Change and from the UK Economic and Social Research Council (grant RES-518-28-001).

APPENDIX A. INCREASING UK CHILD BENEFIT IN 2001 AND 2013: THE NET EFFECTS

In both 2001 and 2013, the UK Child Benefit was delivered as a universal benefit for all children under the age of 19 in full-time nonadvanced education. In both years there were two rates, one for the oldest child (£15.50 and £20.30 per week, respectively) and one for any other children (£10.35 and £13.40 per week, respectively). As an illustration, we double these values and use EUROMOD to calculate the net budgetary cost after the operation of the rest of the tax and benefit systems, and we also show how the gain per child would vary across the household income distribution.

In 2001, Child Benefit was disregarded by the income tax system but was taken into account for the assessment of Income Support (and income-related Job Seeker's Allowance), Housing Benefit, and Council Tax Benefit, some of the main UK means-tested benefits for working-age people and their families. (The Working Families Tax Credit disregarded Child Benefit.) As the table shows, although the gross cost of the increase in Child Benefit is estimated at £8.85 billion per year, once the reduced entitlements to these benefits are taken into account, the net cost falls to £7.01 billion or 79% of the gross.

In contrast, in the 2013 system, the Child Benefit is disregarded in the assessment of all means-tested payments, but higher-income parents who pay income tax at the 40% (or higher) marginal rate have the value of their Child Benefit included in their tax calculation. Thus, as shown in the table, the cost of the increase in Child Benefit is offset to a small extent by an increase in income tax liabilities. In addition, in 2013, there was a cap on the overall sum of benefits that could be received by families in some circumstances. This would result in some families not receiving all or any of their Child Benefit increase. In 2013, the gross cost of the increase in Child Benefit is estimated at £11.55 billion per year, and once the reduced entitlements to these benefits are taken into account, the net cost falls to £11.14 billion or 96% of the gross.

Gross and net cost of doubling Child Benefit, 2001 and 2013

	2001		2013	
	£ Million per year	% of Gross cost	£ Million per year	% of Gross cost
Child Benefit	*8,850*	*100*	*11,549*	*100*
Income Tax	0	0	−290	−3
Income Support	−1,606	−18	0	0
Housing Benefit	−152	−2	0	0
Council Tax Benefit	−81	−1	0	0
Benefit cap	0	0	−123	−1
Net cost	*7,011*	*79*	*11,136*	*96*

Source: EUROMOD version F6.20, using Family Resources Survey data for 2008–2009, adjusted to 2001 and 2013 prices and incomes.

There are different distributional consequences of these differences between gross and net effects, as shown in Figure 24.A1 below. This shows the average net weekly increase in income per child by decile group of equivalized household income under the 2001 and 2013 policy systems. Under the 2001 system, those in the lower income groups receive less, because some of the additional income is withdrawn as reduced entitlement to the means-tested benefit. (This applies to a lesser extent in the bottom decile group in which families simulated to not take-up their entitlements to means-tested benefits are mainly located.) In 2013, however, it is children in higher income households who benefit to a lesser extent, due to the clawback through income tax (the effect of the benefit cap is small and concentrated in the lower-middle of the distribution).

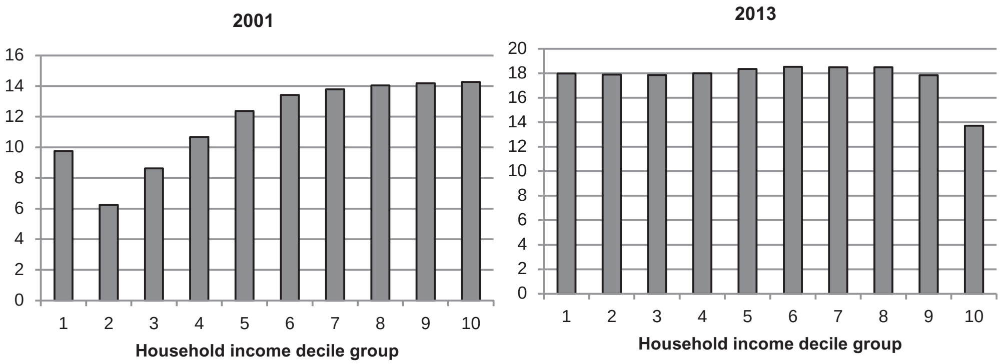


Figure 24.A1 Doubling Child Benefit in the UK: Average net gain per child in £ per week. *Notes: Deciles are based on equivalized household disposable income in the respective years and are constructed using the modified OECD equivalence scale to adjust incomes for differences in household size and composition. The lowest income group is labeled "1" and the highest "10." Source: EUROMOD version F6.20, using Family Resources Survey data for 2008–2009, adjusted to 2001 and 2013 prices and incomes.*

The point of this illustration is to demonstrate how the interactions matter and need to be understood when designing policy scenarios. Similarly, the policy analyst needs to account for the interactions in order to understand the effects of policy changes. If policy-makers wanted to double the payment made to all children in 2001, they would have needed to increase child amounts within the other benefits as well as in Child Benefit. On the other hand, if the goal had been to reduce the number of families subject to means tests (without anyone losing), then the illustrative reform would have done just that (for example, reducing the number of all households receiving Council Tax Benefit). If the goal in 2013 had been to reduce the reach of means-testing, the means-tested payment rates for children would have needed to be reduced at the same time as Child Benefit increase.

APPENDIX B. COMPARISON OF SIMULATED ESTIMATES OF INCOME TAX WITH ADMINISTRATIVE STATISTICS, UK 2010–2011

Here we illustrate the type of validation of simulated income tax that can be carried out using published tables from administrative data of tax revenues. The exercise also suggests ways in which the input micro-data might be adjusted, or not. In this exercise, the input data are the UK Family Resources Survey (FRS) 2009–2010 updated to 2010–2011 incomes and prices.

Simulated income tax liabilities are compared with statistics on income tax paid by band of taxable income, published by the HM Revenue and Customs (HMRC, Table 3.3). The first point to note is that the tax paid in any year may not match the liability for tax on income earned in that year, because of adjustments carried over from previous years.

The first row in the top panel of the table below shows the ratio of microsimulation model (EUROMOD) estimates to those of HMRC in three dimensions: the number of taxpayers (defined as individuals with positive taxable income before deduction of any personal allowances), their total taxable income (before deduction of allowances), and the total tax liability/revenue. The number of taxpayers is underestimated by 7% and taxable income by more: 13%. Also shown are the ratios for the lowest taxable income group (under £10,000 per year) and highest income group (over £150,000).[35]

[35] The HMRC statistics provide more detail for top incomes (the top two groups being £0.5 million to £1 million and £1 million+). However, although the overall sample size of the FRS is large by international standards (31,644 individuals in 2009/10), the numbers with very high incomes are too small to analyze. There are 99 observations with taxable incomes in excess of £150,000, including 13 with more than £500,000 and just 2 with more than £1 million (after adjustment to 2010–2011 income levels). As explained in the main text, this is partly due to underrepresentation of, or underreporting by, people with very high incomes in the survey. However, even if their incomes were properly represented, based on the HMRC statistics, there would still be fewer than 30 observations with incomes above £500,000 in a sample of this size.

Ratio of EUROMOD estimates to HMRC statistics

	Taxpayers	Taxable income	Tax revenue
EUROMOD			
All	0.93	0.87	0.85
Taxable income <£10K	0.99	0.98	0.76
Taxable income £150K+	0.46	0.42	0.46
EUROMOD with proportional adjustment to number of taxpayers (adj1)			
All	1.00	0.93	0.91
Taxable income <£10K	1.06	1.05	0.82
Taxable income £150K+	0.50	0.45	0.49
EUROMOD with adjustment to number of taxpayers by income band (adj2)			
All	1.00	0.98	1.05
Taxable income <£10K	1.00	1.00	0.79
Taxable income £150K+	1.00	0.89	0.98

Notes: EUROMOD-adj1 makes a proportional adjustment to the number of taxpayers, to match the total number in HMRC. EUROMOD-adj2 adjusts the number of taxpayers in each income group to match those given by HMRC.
Sources: EUROMOD version F6.20, using Family Resources Survey 2009–2010 updated to 2010–2011 incomes. HMRC http://www.hmrc.gov.uk/statistics/income-by-year.htm%202010-11%20Table%203.3.

The low-income group and their taxable income is well-represented by the FRS data, but the tax revenue simulated by EUROMOD is too low by 24%. The difference in tax liability based on current year incomes (EUROMOD calculations) and tax revenue (HMRC) can arise if the latter contains taxes due on higher previous-year incomes, for example, by the self-employed. The symmetrical effect (revenue smaller than liability, as incomes rise), which in general is equally likely, is not observable because it is distributed throughout the rest of the distribution of taxable income.

The high-income group of taxpayers and their income and tax paid are all underrepresented by more than 50%. This is consistent with either or both the underrepresentation of very high income-earners in the survey or the underreporting of high incomes.

More detail of the pattern of tax revenue by range of taxable income can be seen in Figure 24.B1 below. The black bars show the HMRC estimates, and the white bars show the estimates using EUROMOD simulations. EUROMOD shows some shortfall in most income ranges, but the effect is concentrated in the highest-income group.
The overall shortfall in taxpayers might be explained by the underreporting of income by the whole distribution or parts of it, or by the fact that a proportion of UK income-tax payers are not resident in UK households and hence not captured by the survey data.[36]

[36] We do not pursue this second possibility further.

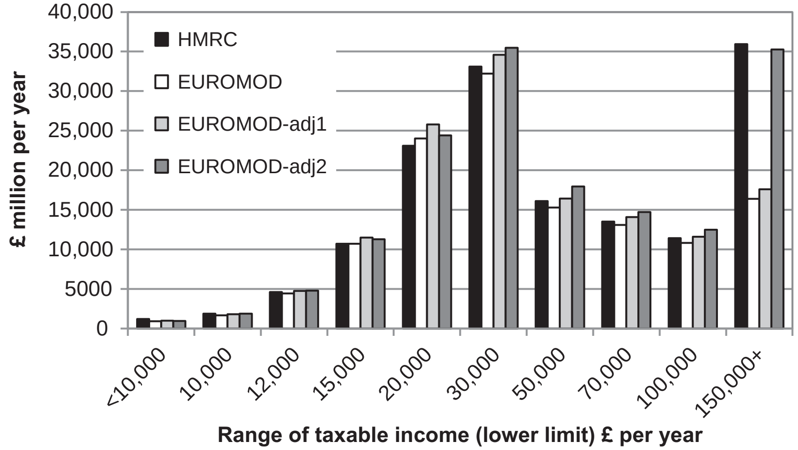


Figure 24.B1 Estimates of income tax revenue by range of taxable income 2010–2011. *Notes: EUROMOD-adj1 makes a proportional adjustment to the number of taxpayers in each income group to match HMRC. EUROMOD-adj2 adjusts the number of taxpayers in each income group to match those given by HMRC. Sources: EUROMOD version F6.20, using Family Resources Survey 2009–2010 updated to 2010–2011 incomes. HMRC http://www.hmrc.gov.uk/statistics/income-by-year.htm%202010-11%20Table%203.3*.

We carry out two adjustments to the comparison of EUROMOD and HMRC statistics (not to the micro-data) in order to explore these possibilities. First, (adj1) we make a proportional adjustment to the three statistics reported in the table such that the total number of taxpayers matches that given by HMRC (i.e., an increase of 7% in the EUROMOD statistic). The middle panel in the table shows that this has the effect of reducing the overall shortfall in taxable income and tax revenue by almost half but does little to rectify the shortfalls in the top income group. The effect on tax revenue in each income group is shown by the pale gray bars in Figure 24.B1. This suggests that a general tendency to underreport is part of the explanation for the shortfall in tax revenue, but it is not the whole story.

A second adjustment accounts for the shortfall in taxpayers within each income group. The effect of this is shown by the darker-gray bar in the figure and the bottom panel in the table. The overall shortfall in taxable income all but disappears, but tax revenue is overestimated by 5%. The shortfall in taxable income in the top income group is greatly reduced to 11%, but tax revenue almost matches that in HMRC statistics for this group. The remaining shortfall in taxable income at the top suggests that there is a problem of lack of response by high-income earners. The relative overestimation of tax revenue suggests that the simulation of tax liability is missing the effect of some tax reliefs and deductions that cannot be simulated due to lack of information in the data. Tax evasion is also a possible explanation. From the figure, it seems that these effects are more important at higher incomes.

To summarize, the validation exercise of the simulated income tax in the UK component of EUROMOD provides some useful insights that should be considered in interpreting microsimulation results for other countries as well, although, of course, the specific explanations may differ, and additional factors may be present. The UK FRS data appear to (a) underrepresent people with high taxable incomes and (b) underreport taxable incomes to some degree, across the whole distribution, although, in countries characterized by large tax evasion, the taxable income reported in the survey can be larger than the one reported in the tax revenue statistics. Simulated tax liabilities do not match tax revenue statistics because of between-year adjustments. The income tax simulations do not account for all reliefs and deductions. Nor do they account for tax evasion, and this may lead to overestimation of tax payments, particularly towards the top of the distribution.

REFERENCES

Aaberge, R., Colombino, U., 2013. Designing optimal taxes with a microeconometric model of household labour supply. Scand. J. Econ. 115 (2), 449–475.
Aaberge, R., Dagsvik, J.K., Strøm, S., 1995. Labour supply responses and welfare effects of tax reforms. Scand. J. Econ. 97 (4), 635–659.
Aaberge, R., Colombino, U., Strøm, S., Wennemo, T., 2000. Joint labour supply of married couples: efficiency and distribution effects of tax and labour market reforms. In: Mitton, L., Sutherland, H., Weeks, M. (Eds.), Microsimulation Modelling for Policy Analysis: Challenges and Innovations. Cambridge University Press, Cambridge, pp. 269–291 (Chapter 13).
Aaberge, R., Colombino, U., Holmoy, E., Strøm, B., Wennemo, T., 2007. Population ageing and fiscal sustainability: integrating detailed labour supply models with CGE models. In: Harding, A., Gupta, A. (Eds.), Modelling Our Future: Population Ageing, Social Security and Taxation. International Symposia in Economic Theory and Econometrics, vol. 15. Elsevier, Amsterdam, pp. 259–290 (Chapter 10).
Aaberge, R., Bhuller, M., Langørgen, A., Mogstad, M., 2010a. The distributional impact of public services when needs differ. J. Public Econ. 94 (9–10), 549–562.
Aaberge, R., Langørgen, A., Lindgren, P., 2010b. The impact of basic public services on the distribution of income in European countries. In: Atkinson, A.B., Marlier, E. (Eds.), Income and Living Conditions in Europe. Eurostat Statistical Books, Publications Office of the European Union, Luxembourg, pp. 329–344 (Chapter 15).
Abramovsky, L., Attanasio, O., Phillips, D., 2012. Demand responses to changes in consumer prices in Mexico: lessons for policy and an application to the 2010 Mexican tax reforms. Conference paper presented at the 2012 Annual Conference of the Royal Economic Society, University of Cambridge.
Adam, S., Brewer, M., Shephard, A., 2006a. Financial work incentives in Britain: comparisons over time and between family types, IFS Working Papers 20, The Institute for Fiscal Studies.
Adam, S., Brewer, M., Shephard, A., 2006b. The Poverty Trade-off: Work Incentives and Income Redistribution in Britain. The Joseph Rowntree Foundation, The Policy Press, Bristol.
Alm, J., 2012. Measuring, explaining, and controlling tax evasion: lessons from theory, experiments, and field studies. Int. Tax Public Financ. 19, 54–77.
Alm, J., McClelland, G.H., Schulze, W.D., 1992. Why do people pay taxes? J. Public Econ. 48 (1), 21–38.
Alm, J., Deskins, J., McKee, M., 2009. Do individuals comply on income not reported by their employer? Public Financ. Rev. 37 (2), 120–141.
Alm, J., Cherry, T.L., Jones, M., McKee, M., 2012. Social programs as positive inducements for tax participation. J. Econ. Behav. Organ. 84 (1), 85–96.
Anderson, P.M., Meyer, B.D., 1997. Unemployment insurance takeup rates and the after-tax value of benefits. Q. J. Econ. 112 (3), 913–937.

Anderson, B., Agostini, P.D., Lawson, T., 2014. Estimating the small area effects of austerity measures in the UK. In: Dekkers, G., Keegan, M., O'Donoghue, C. (Eds.), New Pathways in Microsimulation. Ashgate, Farnham, pp. 11–28 (Chapter 2).

Andreoni, J., Erard, B., Feinstein, J., 1998. Tax compliance. J. Econ. Lit. 36 (2), 818–860.

Atkinson, A.B., 2009. An enlarged role for tax-benefit models. In: Lelkes, O., Sutherland, H. (Eds.), Tax and Benefit Policies in the Enlarged Europe: Assessing the Impact with Microsimulation Models. Ashgate, Vienna, pp. 33–46 (Chapter 2).

Atkinson, A.B., Marlier, E., 2010. Living conditions in Europe and the Europe 2020 agenda. In: Atkinson, A.B., Marlier, E. (Eds.), Income and Living Conditions in Europe. Eurostat Statistical Books, Publications Office of the European Union, Luxembourg, pp. 21–35 (Chapter 1).

Atkinson, A.B., Sutherland, H., 1989. Scaling the "poverty mountain": methods to extend incentives to all workers. In: Bowen, A., Mayhew, K. (Eds.), Improving Incentives for the Low Paid. NEDO, Macmillan, London.

Atkinson, A.B., King, M.A., Sutherland, H., 1983. The analysis of personal taxation and social security. Natl. Inst. Econ. Rev. 103, 63–74.

Atkinson, A.B., Bourguignon, F., Chiappori, P.-A., 1988. What do we learn about tax reform from international comparisons? France and Britain. Eur. Econ. Rev. 32 (2–3), 343–352.

Atkinson, A.B., Bourguignon, F., O'Donoghue, C., Sutherland, H., Utili, F., 2002. Microsimulation of social policy in the European Union: case study of a European minimum pension. Economica 69, 229–243.

Atta-Darkua, V., Barnard, A., 2010. Distributional effects of direct taxes and social transfers (cash benefits). In: Atkinson, A.B., Marlier, E. (Eds.), Income and Living Conditions in Europe. Eurostat Statistical Books, Publications Office of the European Union, Luxembourg, pp. 345–368 (Chapter 16).

Auerbach, A.J., Feenberg, D., 2000. The significance of federal taxes as automatic stabilizers. J. Econ. Perspect. 14 (3), 37–56.

Avram, S., Figari, F., Leventi, C., Levy, H., Navicke, J., Matsaganis, M., Militaru, E., Paulus, A., Rastrigina, O., Sutherland, H., 2013. The distributional effects of fiscal consolidation in nine EU countries. EUROMOD Working Paper EM2/13, University of Essex, Colchester.

Azzolini, D., Bazzoli, M., De Poli, S., Fiorio, C., Poy, S., 2014. TREMOD: A Microsimulation Model for the Province of Trento (Italy), EUROMOD Working Paper EM15/14. University of Essex, Colchester.

Banbura, M., Giannone, D., Reichlin, L., 2011. Nowcasting. In: Clements, M.P., Hendry, D.F. (Eds.), The Oxford Handbook of Economic Forecasting. Oxford University Press, Oxford.

Bargain, O. (Ed.), 2007. Micro-Simulation in Action: Policy Analysis in Europe Using EUROMOD. Research in Labor Economics, vol. 25. Elsevier, Oxford.

Bargain, O., 2012a. Decomposition analysis of distributive policies using behavioural simulations. Int. Tax Public Financ. 19 (5), 708–731.

Bargain, O., 2012b. The distributional effects of tax-benefit policies under New Labour: a decomposition approach. Oxf. Bull. Econ. Stat. 74 (6), 856–874.

Bargain, O., Callan, T., 2010. Analysing the effects of tax-benefit reforms on income distribution: a decomposition approach. J. Econ. Inequal. 8 (1), 1–21.

Bargain, O., Orsini, K., 2007. Beans for breakfast? How portable is the British workfare model? In: Bargain, O. (Ed.), Micro-Simulation in Action: Policy Analysis in Europe Using EUROMOD. Research in Labor Economics, vol. 25. Elsevier, Oxford, pp. 165–198.

Bargain, O., Immervoll, H., Viitamäki, H., 2012. No claim, no pain. Measuring the non-take-up of social assistance using register data. J. Econ. Inequal. 10 (3), 375–395.

Bargain, O., Dolls, M., Fuest, C., Neumann, D., Peichl, A., Pestel, N., Siegloch, S., 2013a. Fiscal union in Europe? Redistributive and stabilizing effects of a European tax-benefit system and fiscal equalization mechanism. Econ. Policy 28 (75), 375–422.

Bargain, O., Dolls, M., Immervoll, H., Neumann, D., Peichl, A., Pestel, N., Siegloch, S., 2013b. Partisan tax policy and income inequality in the U.S., 1979–2007, IZA Discussion Paper 7190.

Bargain, O., Orsini, K., Peichl, A., 2014. Comparing labor supply elasticities in Europe and the US: new results. J. Hum. Resour. 49 (3), 723–838.

Benedek, D., Lelkes, O., 2011. The distributional implications of income under-reporting in Hungary. Fisc. Stud. 32 (4), 539–560.
Bennett, F., Sutherland, H., 2011. The importance of independent income: understanding the role of non-means-tested earnings replacement benefits. ISER Working Paper 2011-09, University of Essex, Colchester.
Betson, D., Greenberg, D., Kasten, R., 1982. A simulation analysis of the economic efficiency and distribution effects of alternative program structures: the negative income tax versus the credit income tax. In: Garfinkel, I. (Ed.), *Income Tested Transfer Programs: The Case For and Against*. Academic Press, New York, pp. 175–203 (Chapter 6).
Betti, G., Donatiello, G., Verma, V., 2011. The Siena Microsimulation Model (SM2) for net-gross conversion of EU-SILC income variables. Int. J. Microsimulation 4 (1), 35–53.
Bingley, P., Walker, I., 1997. The labour supply, unemployment and participation of lone mothers in in-work transfer programmes. Econ. J. 107 (444), 1375–1390.
Bingley, P., Walker, I., 2001. Housing subsidies and work incentives in Great Britain. Econ. J. 111 (471), C86–C103.
Bitler, M.P., Currie, J., Scholz, J.K., 2003. WIC eligibility and participation. J. Hum. Resour. 38, 1139–1179.
Blank, R.M., Ruggles, P., 1996. When do women use aid to families with dependent children and food stamps? The dynamics of eligibility versus participation. J. Hum. Resour. 31 (1), 57–89.
Blazevski, N.M., Petreski, M., Petreska, D., 2013. Increasing labour market activity of the poor and females: Let's make work pay in Macedonia. EUROMOD Working Paper EM16/13, University of Essex, Colchester.
Blundell, R., 2006. Earned income tax credit policies: impact and optimality. The Adam Smith Lecture, 2005. Labour Econ. 13 (4), 423–443.
Blundell, R., 2012. Tax policy reform: the role of empirical evidence. J. Eur. Econ. Assoc. 10 (1), 43–77.
Blundell, R., Shepard, A., 2012. Employment, hours of work and the optimal taxation of low income families. Rev. Econ. Stud. 79, 481–510.
Blundell, R., Fry, V., Walker, I., 1988. Modelling the take-up of means-tested benefits: the case of housing benefits in the United Kingdom. Econ. J. 98 (390), 58–74.
Blundell, R., Chiappori, P., Magnac, T., Meghir, C., 2007. Collective labour supply: heterogeneity and nonparticipation. Rev. Econ. Stud. 74, 417–445.
Boadway, R., Wildasin, D., 1995. Taxation and savings: a survey. Fisc. Stud. 15 (3), 19–63.
Borella, M., Coda Moscarola, F., 2010. Microsimulation of pension reforms: behavioural versus non behavioural approach. J. Pension Econ. Financ. 9 (4), 583–607.
Bourguignon, F., Bussolo, M., 2013. Income distribution in computable general equilibrium modelling. In: Dixon, P.B., Jorgenson, D.W. (Eds.), Handbook of Computable General Equilibrium Modelling. vol. 1B. Elsevier, Amsterdam, pp. 1383–1437 (Chapter 21).
Bourguignon, F., Spadaro, A., 2006. Microsimulation as a tool for evaluating redistribution policies. J. Econ. Inequal. 4 (1), 77–106.
Bourguignon, F., Chiappori, P.A., Hugounenq, R., 1993. Exploring the distribution and incentive effects of tax harmonization. In: Heimler, A., Meulders, D. (Eds.), Empirical Approaches to Fiscal Policy Modelling. Chapman and Hall, London, pp. 235–250 (Chapter 11).
Bourguignon, F., Robilliard, A.S., Robinson, S., 2005. Representative versus real households in the macroeconomic modelling of inequality. In: Kehoe, T.J., Srinivasan, T.N., Whalley, J. (Eds.), Frontiers in Applied General Equilibrium Modelling. Cambridge University Press, Cambridge, pp. 219–254 (Chapter 10).
Bozio, A., Fabre, B., Goupille, J., Lafféter, Q., 2012. Le modéle de micro-simulation TAXIPP—version 0.2. Institut des Politiques Publique, Paris.
Brandolini, A., D'Amuri, F., Faiella, I., 2013. Country case study—Italy. In: Jenkins, S.P., Brandolini, A., Micklewright, J., Nolan, B. (Eds.), The Great Recession and the Distribution of Household Income. Oxford University Press, Oxford, pp. 130–152 (Chapter 7).
Brewer, M., O'Dea, C., 2012. Measuring living standards with income and consumption: evidence from the UK. ISER Working Paper 2012-05, University of Essex, Colchester.

Brewer, M., Francesconi, M., Gregg, P., Grogger, J., 2009. In-work benefit reform in a cross-national perspective—introduction. Econ. J. 119 (535), F1–F14.

Brewer, M., Saez, E., Shephard, A., 2010. Means-testing and tax rates on earnings. In: Mirrlees, J., Adam, S., Besley, T., Blundell, R., Bond, S., Chote, R., Gammie, M., Johnson, P., Myles, G., Poterba, J. (Eds.), Dimensions of Tax Design: The Mirrlees Review. Oxford University Press, Oxford, pp. 90–173, (Chapter 2).

Brewer, M., Browne, J., Joyce, R., 2011. Child and working-age poverty from 2010 to 2020. IFS Commentary C121, The Institute for Fiscal Studies, London.

Brewer, M., Browne, J., Hood, A., Joyce, R., Sibieta, L., 2013. The short- and medium-term impacts of the recession on the UK income distribution. Fisc. Stud. 34 (2), 179–201.

Brown, L.J., Harris, A., Picton, M., Thurecht, L., Yap, M., Harding, A., Dixon, P.B., Richardson, J., 2009. Linking microsimulation and macro-economic models to estimate the economic impact of chronic disease prevention. In: Zaidi, A., Harding, A., Williamson, P. (Eds.), New Frontiers in Microsimulation Modelling. Ashgate, Vienna, pp. 527–555 (Chapter 20).

Bruckmeier, K., Wiemers, J., 2012. A new targeting: a new take-up? Non-take-up of social assistance in Germany after social policy reforms. Empir. Econ. 43 (2), 565–580.

Caldwell, S.B., 1990. Static, Dynamic and Mixed Microsimulation. Department of Sociology, Cornell University, Ithaca.

Callan, T., Sutherland, H., 1997. The impact of comparable policies in European countries: Microsimulation approaches. Eur. Econ. Rev. 41 (3-5), 627–633.

Callan, T., Nolan, B., Walsh, J., 1999. Income tax and social welfare policies. In: Budget Perspectives 1999. ESRI, Dublin.

Callan, T., Coleman, K., Walsh, J.R., 2007. Assessing the impact of tax-transfer policy changes on poverty: methodological issues and some European evidence. In: Bargain, O. (Ed.), Micro-Simulation in Action: Policy Analysis in Europe Using EUROMOD. Research in Labor Economics, vol. 25. Elsevier, Oxford.

Cameron, G., Ezzeddin, R., 2000. Assessing the direct and indirect effects of social policy: integrating input-output and tax microsimulation models at Statistics Canada. In: Mitton, L., Sutherland, H., Weeks, M. (Eds.), Microsimulation Modelling for Policy Analysis: Challenges and Innovations. Cambridge University Press, Cambridge, pp. 42–65 (Chapter 3).

Cantó, O., Adiego, M., Ayala, L., Levy, H., Paniagua, M., 2014. Going regional: The effectiveness of different tax-benefit policies in combating child poverty in Spain. In: Dekkers, G., Keegan, M., O'Donoghue, C. (Eds.), New Pathways in Microsimulation. Ashgate, Farnham, pp. 183–202 (Chapter 12).

Capéau, B., Decoster, A., Phillips, D., 2014. Micro-simulation models of consumption and indirect taxation. In: O'Donoghue, C. (Ed.), Handbook of Microsimulation Modelling. Emerald, Bingley (forthcoming).

Ceriani, L., Fiorio, C.V., Gigliarano, C., 2013. The importance of choosing the data set for tax-benefit analysis. Int. J. Microsimulation 6 (1), 86–121.

Chan, M.K., 2013. A dynamic model of welfare reform. Econometrica 81 (3), 941–1001.

Clark, T., Leicester, A., 2004. Inequality and two decades of British tax and benefit reforms. Fisc. Stud. 25 (2), 129–158.

Cleveland, R.W., 2005. Alternative income estimates in the United States: 2003, Current Population Reports P60-228, U.S. Census Bureau.

Clotfelter, C., 1983. Tax evasion and tax rates: an analysis of individual returns. Rev. Econ. Stat. 65 (3), 363–373.

Colombino, U., 2013. A new equilibrium simulation procedure with discrete choice models. Int. J. Microsimulation 6 (3), 25–49.

Cowell, F.A., 2000. Measurement of inequality. In: Atkinson, A.B., Bourguignon, F. (Eds.), Handbook of Income Distribution. vol. 1. Elsevier, Amsterdam, pp. 87–166 (Chapter 2).

Creedy, J., 1999a. Lifetime versus annual income distribution. In: Silber, J. (Ed.), Handbook on Income Inequality Measurement. Kluwer Academic Publishing, Dordrecht, pp. 513–533 (Chapter 17).

Creedy, J., 1999b. Modelling Indirect Taxes and Tax Reform. Edward Elgar, Northampton.

Creedy, J., 2004. Survey reweighting for tax microsimulation modelling. Res. Econ. Inequal. 12, 229–249.

Creedy, J., Duncan, A., 2002. Behavioural microsimulation with labour supply responses. J. Econ. Surv. 16 (1), 1–39.
Creedy, J., Hérault, N., 2011. Decomposing inequality and social welfare changes: the use of alternative welfare metrics, Melbourne Institute Working Paper 8/11, University of Melbourne.
Creedy, J., Kalb, G., 2005. Discrete hours labour supply modelling: specification, estimation and simulation. J. Econ. Surv. 19 (5), 697–734.
Creedy, J., Kalb, G., Kew, H., 2007. Confidence intervals for policy reforms in behavioural tax microsimulation modelling. Bull. Econ. Res. 59 (1), 37–65.
Currie, J., 2004. The take up of social benefits, NBER Working Paper 10488.
Daponte, B.O., Sanders, S., Taylor, L., 1999. Why do low-income households not use Food Stamps? Evidence from an experiment. J. Hum. Resour. 34 (3), 612–628.
De Lathouwer, L., 1996. A case study of unemployment scheme for Belgium and the Netherlands. In: Harding, A. (Ed.), Microsimulation and Public Policy. Contributions to Economic Analysis, vol. 232. North-Holland, Amsterdam, pp. 69–92 (Chapter 4).
de Mooij, R., Keen, M., 2013. 'Fiscal devaluation' and fiscal consolidation: the VAT in troubled times. In: Alesina, A., Giavazzi, F. (Eds.), Fiscal Policy After the Financial Crisis. University of Chicago Press, Chicago, pp. 443–485 (Chapter 11).
de Vos, K., Zaidi, A., 1996. The use of microsimulation to update poverty statistics based on household budget surveys: a pilot study for the UK. In: Harding, A. (Ed.), Microsimulation and Public Policy. Contributions to Economic Analysis, vol. 232. North-Holland, Amsterdam, pp. 111–128 (Chapter 6).
Decancq, K., Decoster, A., Spiritus, K., Verbist, G., 2012. MEFISTO: a new micro-simulation model for Flanders, FLEMOSI Discussion Paper 14.
Decoster, A., Van Camp, G., 2000. The unit of analysis in microsimulation models for personal income taxes: fiscal unit or household? In: Mitton, L., Sutherland, H., Weeks, M. (Eds.), Microsimulation Modelling for Policy Analysis: Challenges and Innovations. Cambridge University Press, Cambridge, pp. 15–41 (Chapter 2).
Decoster, A., Van Camp, G., 2001. Redistributive effects of the shift from personal income taxes to indirect taxes: Belgium 1988–93. Fisc. Stud. 22 (1), 79–106.
Decoster, A., Loughrey, J., O'Donoghue, C., Verwerft, D., 2010. How regressive are indirect taxes? A microsimulation analysis for five European countries. J. Policy Anal. Manage. 29 (2), 326–350.
Decoster, A., Loughrey, J., O'Donoghue, C., Verwerft, D., 2011. Microsimulation of indirect taxes. Int. J. Microsimulation 4 (2), 41–56.
Dekkers, G., Buslei, H., Cozzolino, M., Desmet, R., Geyer, J., Hofmann, D., Raitano, M., Steiner, V., Tanda, P., Tedeschi, S., Verschueren, F., 2010. The flip side of the coin: the consequences of the European budgetary projections on the adequacy of social security pensions. Eur. J. Soc. Secur. 12 (2), 94–121.
Dekkers, G., Keegan, M., O'Donoghue, C. (Eds.), 2014. New Pathways in Microsimulation. Ashgate, Farnham.
Doerrenberg, P., Duncan, D., 2013. Distributional implications of tax evasion: evidence from the lab. Public Financ. Rev. (forthcoming).
Dolls, M., Fuest, C., Peichl, A., 2012. Automatic stabilizers and economic crisis: US vs. Europe. J. Public Econ. 96 (3–4), 279–294.
Dorsett, R., Heady, C., 1991. The take-up of means-tested benefits by working families with children. Fisc. Stud. 12 (4), 22–32.
Dowling, R., Skabardonis, J., Halkias, J., McHale, G., Zammit, G., 2004. Guidelines for calibration of microsimulation models: framework and applications. Transport. Res. Rec. 1876 (1), 1–9.
Duclos, J.-Y., 1995. Modelling the take-up of state support. J. Public Econ. 58 (3), 391–415.
Duclos, J.-Y., 1997. Estimating and testing a model of welfare participation: the case of supplementary benefits in Britain. Economica 64 (253), 81–100.
DWP, 2013. Households below average income: an analysis of the income distribution 1994/95–2011/12. Department for Work and Pensions, London.
Elbers, C., Lanjouw, J., Lanjouw, P., 2003. Micro-level estimation of poverty and inequality. Econometrica 71 (1), 355–364.
Elffers, H., Robben, H.S., Hessing, D.J., 1992. On measuring tax evasion. J. Econ. Psychol. 13 (4), 545–567.
Erard, B., 1993. Taxation with representation: an analysis of the role of tax practitioners in tax compliance. J. Public Econ. 52 (2), 163–197.

Erard, B., 1997. Self-selection with measurement errors. A microeconometric analysis of the decision to seek tax assistance and its implications for tax compliance. J. Econ. 81 (2), 319–356.
Erard, B., Ho, C.-C., 2001. Searching for ghosts: who are the nonfilers and how much tax do they owe? J. Public Econ. 81 (1), 25–50.
European Commission, 2013. Study on the impacts of fiscal devaluation. Number 36 in Taxation Papers, Publications Office of the European Union, Luxembourg.
Feenberg, D.R., Coutts, E., 1993. An introduction to the TAXSIM model. J. Policy Anal. Manage. 12 (1), 189–194.
Feinstein, J.S., 1991. An econometric analysis of income tax evasion and its detection. RAND J. Econ. 22 (1), 14–35.
Feldstein, M.S., Feenberg, D.R., 1983. Alternative tax rules and personal saving incentives: microeconomic data and behavioral simulations. In: Feldstein, M.S. (Ed.), Behavioral Simulation Methods in Tax Policy Analysis. Chicago, London.
Fernández Salgado, M., Figari, F., Sutherland, H., Tumino, A., 2014. Welfare compensation for unemployment in the Great Recession. Rev. Income Wealth 60, S177–S204.
Figari, F., 2010. Can in-work benefits improve social inclusion in the southern European countries? J. Eur. Soc. Policy 20 (4), 301–315.
Figari, F., Paulus, A., 2013. The distributional effects of taxes and transfers under alternative income concepts: the importance of three 'I's. Public Financ. Rev. (forthcoming).
Figari, F., Levy, H., Sutherland, H., 2007. Using the EU-SILC for policy simulation: prospects, some limitations and suggestions. In: Comparative EU Statistics on Income and Living Conditions: Issues and Challenges. Eurostat Methodologies and Working Papers, Office for Official Publications of the European Communities, Luxembourg, pp. 345–373.
Figari, F., Immervoll, H., Levy, H., Sutherland, H., 2011a. Inequalities within couples in Europe: market incomes and the role of taxes and benefits. East. Econ. J. 37, 344–366.
Figari, F., Paulus, A., Sutherland, H., 2011b. Measuring the size and impact of public cash support for children in cross-national perspective. Soc. Sci. Comput. Rev. 29 (1), 85–102.
Figari, F., Salvatori, A., Sutherland, H., 2011c. Economic downturn and stress testing European welfare systems. In: Immervoll, H., Peichl, A., Tatsiramos, K. (Eds.), Who Loses in the Downturn? Economic Crisis, Employment and Income Distribution. Research in Labor Economics, vol. 32. Emerald Group Publishing Limited, Bingley, pp. 257–286.
Figari, F., Iacovou, M., Skew, A.J., Sutherland, H., 2012a. Approximations to the truth: comparing survey and microsimulation approaches to measuring income for social indicators. Soc. Indic. Res. 105 (3), 387–407.
Figari, F., Paulus, A., Sutherland, H., Tsakloglou, P., Verbist, G., Zantomio, F., 2012b. Taxing home ownership: distributional effects of including net imputed rent in taxable income. EUROMOD Working Paper EM4/12, University of Essex, Colchester.
Flood, L., 2007. Can we afford the future? An evaluation of the new Swedish pension system. In: Harding, A., Gupta, A. (Eds.), Modelling Our Future: Population Ageing, Social Security and Taxation. International Symposia in Economic Theory and Econometrics, vol. 15. Elsevier, Amsterdam, pp. 33–54 (Chapter 2).
Forest, A., Sheffrin, S.M., 2002. Complexity and compliance: an empirical investigation. Natl. Tax J. 55 (1), 75–88.
Fortin, N., Lemieux, T., Firpo, S., 2011. Decomposition methods in economics. In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics. vol. 4, Part A. Elsevier, Amsterdam, pp. 1–102 (Chapter 1).
Frick, J.R., Grabka, M.M., Smeeding, T.M., Tsakloglou, P., 2010. Distributional effects of imputed rents in five European countries. J. Hous. Econ. 19 (3), 167–179.
Fuest, C., Niehues, J., Peichl, A., 2010. The redistributive effects of tax benefit systems in the enlarged EU. Public Financ. Rev. 38 (4), 473–500.
Fullerton, D., Metcalf, G.E., 2002. Tax incidence. In: Auerbach, A.J., Feldstein, M. (Eds.), Handbook of Public Economics. vol. 4. Elsevier, Amsterdam, pp. 1787–1872 (Chapter 26).
Goedemé, T., Van den Bosch, K., Salanauskaite, L., Verbist, G., 2013. Testing the statistical significance of microsimulation results: a plea. Int. J. Microsimulation 6 (3), 50–77.

Gomulka, J., 1992. Grossing up revisited. In: Hancock, R., Sutherland, H. (Eds.), Microsimulation Models for Public Policy Analysis: New Frontiers. London School of Economics, London, pp. 121–132 (Chapter 6).
Gupta, A., Harding, A. (Eds.), 2007. Modelling Our Future: Population Ageing, Health and Aged Care. International Symposia in Economic Theory and Econometrics, vol. 16. Elsevier, Amsterdam.
Gupta, A., Kapur, V. (Eds.), 2000. Microsimulation in Government Policy and Forecasting. Contributions to Economic Analysis, vol. 247. North-Holland, Amsterdam.
Haan, P., 2010. A multi-state model of state dependence in labour supply. Labour Econ. 17 (2), 323–335.
Haider, S.J., Jacknowitz, A., Schoeni, R.F., 2003. Food stamps and the elderly: why is participation so low? J. Hum. Resour. 38, 1080–1111.
Hancock, R., 2000. Charging for care in later life: an exercise in dynamic microsimulation. In: Mitton, L., Sutherland, H., Weeks, M. (Eds.), Microsimulation Modelling for Policy Analysis: Challenges and Innovations. Cambridge University Press, Cambridge, pp. 226–237 (Chapter 10).
Hancock, R., Barker, G., 2005. The quality of social security benefit data in the British Family Resources Survey: implications for investigating income support take-up by pensioners. J. R. Stat. Soc. Ser. A Stat. Soc. 168 (1), 63–82.
Hancock, R., Pudney, S., 2014. Assessing the distributional impact of reforms to disability benefits for older people in the UK: implications of alternative measures of income and disability costs. Ageing Soc. 34 (2), 232–257.
Hancock, R., Pudney, S., Barker, G., Hernandez, M., Sutherland, H., 2004. The take-up of multiple means-tested benefits by British pensioners: evidence from the Family Resources Survey. Fisc. Stud. 25 (3), 279–303.
Hancock, R., Malley, J., Wittenberg, R., Morciano, M., Pickard, L., King, D., Comas-Herrera, A., 2013. The role of care home fees in the public costs and distributional effects of potential reforms to care home funding for older people in England. Health Econ. Policy Law 8 (1), 47–73.
Harding, A., 1993. Lifetime Income Distribution and Redistribution. Application of a Microsimulation Model. Contributions to Economic Analysis, vol. 221. North-Holland, Amsterdam.
Harding, A., 1996a. Introduction and overview. In: Harding, A. (Ed.), Microsimulation and Public Policy. Contributions to Economic Analysis, vol. 232. North-Holland, Amsterdam, pp. 1–22 (Chapter 1).
Harding, A. (Ed.), 1996b. Microsimulation and Public Policy.Contributions to Economic Analysis, vol. 232. North-Holland, Amsterdam.
Harding, A., Gupta, A. (Eds.), 2007. Modelling Our Future: Population Ageing, Social Security and Taxation. In: International Symposia in Economic Theory and Econometrics, vol. 15. Elsevier, Amsterdam.
Hernandez, M., Pudney, S., 2007. Measurement error in models of welfare participation. J. Public Econ. 91 (1–2), 327–341.
Hernandez, M., Pudney, S., Hancock, R., 2007. The welfare cost of means-testing: pensioner participation in income support. J. Appl. Econ. 22 (3), 581–598.
Hernanz, V., Malherbet, F., Pellizzari, M., 2004. Take-up of welfare benefits in OECD countries: a review of the evidence, Social, Employment and Migration Working Papers 17, OECD, Paris.
HM Treasury, 2013. Budget 2013: policy costings, London.
HMRC, 2012. The Exchequer effect of the 50 per cent additional rate of income tax. HM Revenue & Customs, London.
Hérault, N., 2010. Sequential linking of computable general equilibrium and microsimulation models: comparison of behavioural and reweighting techniques. Int. J. Microsimulation 3 (1), 35–42.
Hungerford, T.L., 2010. The redistributive effect of selected federal transfer and tax provisions. Public Financ. Rev. 38 (4), 450–472.
Hurst, E., Li, G., Pugsley, B., 2014. Are household surveys like tax forms: evidence from income underreporting of the self-employed. Rev. Econ. Stat. 96 (1), 19–33.
Ilmakunnas, S., Pudney, S., 1990. A model of female labour supply in the presence of hours restrictions. J. Public Econ. 41 (2), 183–210.

Immervoll, H., 2004. Average and marginal effective tax rates facing workers in the EU: a micro-level analysis of levels, distributions and driving factors, Social, Employment and Migration Working Papers 19, OECD, Paris.

Immervoll, H., 2005. Falling up the stairs: the effects of 'bracket creep' on household incomes. Rev. Income Wealth 51 (1), 37–62.

Immervoll, H., O'Donoghue, C., 2001. Imputation of gross amounts from net incomes in household surveys: an application using EUROMOD. EUROMOD Working Paper EM1/01, University of Cambridge.

Immervoll, H., O'Donoghue, C., 2004. What difference does a job make? The income consequences of joblessness in Europe. In: Gallie, D. (Ed.), Resisting Marginalisation: Unemployment Experience and Social Policy in the European Union. Oxford University Press, Oxford, pp. 105–139 (Chapter 5).

Immervoll, H., Richardson, L., 2011. Redistribution Policy and Inequality Reduction in OECD Countries: What has Changed in Two Decades?, Social, Employment and Migration Working Papers 122. OECD Publishing, Paris.

Immervoll, H., Levy, H., Lietz, C., Mantovani, D., O'Donoghue, C., Sutherland, H., Verbist, G., 2006a. Household incomes and redistribution in the European Union: quantifying the equalizing properties of taxes and benefits. In: Papadimitriou, D. (Ed.), The Distributional Effects of Government Spending and Taxation. Palgrave Macmillan, Basingstoke, pp. 135–165.

Immervoll, H., Levy, H., Lietz, C., Mantovani, D., Sutherland, H., 2006b. The sensitivity of poverty rates to macro-level changes in the European Union. Camb. J. Econ. 30 (2), 181–199.

Immervoll, H., Kleven, H.J., Kreiner, C.T., Saez, E., 2007. Welfare reform in European countries: a microsimulation analysis. Econ. J. 117 (516), 1–44.

Isaacs, J.B., Healy, O., 2012. The recession's ongoing impact on children, 2012. The Urban Institute.

Jara, H.X., Tumino, A., 2013. Tax-benefit systems, income distribution and work incentives in the European Union. Int. J. Microsimulation 6 (1), 27–62.

Jenkins, S.P., 2011. Changing Fortunes: Income Mobility and Poverty Dynamics in Britain. Oxford University Press, Oxford.

Johns, A., Slemrod, J., 2010. The distribution of income tax noncompliance. Natl. Tax J. 63 (3), 397–418.

Keane, M.P., 2010. Structural vs. atheoretic approaches to econometrics. J. Econ. 156 (1), 3–20.

Keane, M.P., Moffitt, R., 1998. A structural model of multiple welfare program participation and labor supply. Int. Econ. Rev. 39 (3), 553–589.

Keane, C., Callan, T., Savage, M., Walsh, J., Timoney, K., 2013. Identifying policy impacts in the crisis: microsimulation evidence on tax and welfare. J. Stat. Soc. Inquiry Society Ireland 42, 1–14.

Kim, K., Lambert, P.J., 2009. Redistributive effect of U.S. taxes and public transfers, 1994–2004. Public Financ. Rev. 37 (1), 3–26.

King, M.A., 1983. The distribution of gains and losses from changes in the tax treatment of housing. In: Feldstein, M. (Ed.), Behavioural Simulation Methods in Tax Policy Analysis. University of Chicago Press, Chicago, pp. 109–137 (Chapter 4).

Klepper, S., Nagin, D., 1989. The anatomy of tax evasion. J. Law Econ. Organ. 5 (1), 1–24.

Kleven, H.J., Knudsen, M.B., Kreiner, C.T., Pedersen, S., Saez, E., 2011. Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark. Econometrica 79 (3), 651–692.

Klevmarken, N.A., 1997. Behavioral modeling in micro simulation models. A survey. Department of Economics, Uppsala University, Working Paper 31.

Klevmarken, N.A., 2002. Statistical inference in micro-simulation models: incorporating external information. Math. Comput. Simul. 59 (1–3), 255–265.

Klevmarken, N.A., 2008. Dynamic microsimulation for policy analysis: problems and solutions. In: Klevmarken, A., Lindgren, B. (Eds.), Simulating an Ageing Population: A Microsimulation Approach Applied to Sweden. Contributions to Economic Analysis, vol. 285. Emerald Group Publishing Limited, Bingley, pp. 31–53 (Chapter 2).

Kotlikoff, L.J., Rapson, D., 2007. Does it pay, at the margin, to work and save? Measuring effective marginal taxes on Americans' labor supply and saving. Tax Pol. Econ. 21, 83–143.

Lanjouw, P., Marra, M., Nguyen, C., 2013. Vietnam's evolving poverty map: patterns and implications for policy. Policy Research Working Paper 6355, The World Bank, Washington, DC.
Laury, S., Wallace, S., 2005. Confidentiality and taxpayer compliance. Natl. Tax J. 58 (3), 427–438.
Lelkes, O., Sutherland, H. (Eds.), 2009. Tax and Benefit Policies in the Enlarged Europe: Assessing the Impact with Microsimulation Models. Public Policy and Social Welfare, vol. 35. Ashgate, Vienna.
Leventi, C., Matsaganis, M., Flevotomou, M., 2013. Distributional implications of tax evasion and the crisis in Greece. EUROMOD Working Paper EM17/13, University of Essex, Colchester.
Levy, H., Lietz, C., Sutherland, H., 2007a. A guaranteed income for Europe's children? In: Jenkins, S.P., Micklewright, J. (Eds.), Inequality and Poverty Re-Examined. Oxford University Press, Oxford.
Levy, H., Lietz, C., Sutherland, H., 2007b. Swapping policies: alternative tax-benefit strategies to support children in Austria, Spain and the UK. J. Soc. Policy 36, 625–647.
Levy, H., Morawski, L., Myck, M., 2009. Alternative tax-benefit strategies to support children in Poland. In: Lelkes, O., Sutherland, H. (Eds.), Tax and Benefit Policies in the Enlarged Europe: Assessing the Impact with Microsimulation Models. Public Policy and Social Welfare, vol. 35. Ashgate, Vienna, pp. 125–151 (Chapter 6).
Levy, H., Matsaganis, M., Sutherland, H., 2013. Towards a European Union child basic income? Within and between country effects. Int. J. Microsimulation 6 (1), 63–85.
Li, J., O'Donoghoue, C., 2013. A methodological survey of dynamic microsimulation models. Int. J. Microsimulation 6 (2), 3–55.
Liégeois, P., Dekkers, G., 2014. Combining EUROMOD and LIAM tools for the development of dynamic cross-sectional microsimulation models: a snack preview. In: Dekkers, G., Keegan, M., O'Donoghue, C. (Eds.), New Pathways in Microsimulation. Ashgate, Farnham, pp. 203–216 (Chapter 13).
Lymer, S., Brown, L., Harding, A., Yap, M., 2009. Predicting the need for aged care services at the small area level: the CAREMOD Spatial Microsimulation Model. Int. J. Microsimulation 2 (2), 27–42.
Lynn, P., Jäckle, A., Jenkins, S.P., Sala, E., 2012. The impact of questioning method on measurement error in panel survey measures of benefit receipt: evidence from a validation study. J. R. Stat. Soc. Ser. A Stat. Soc. 175 (1), 289–308.
Lyssiotou, P., Pashardes, P., Stengos, T., 2004. Estimates of the black economy based on consumer demand approaches. Econ. J. 114 (497), 622–640.
Mabbett, D., Schelke, W., 2007. Bringing macroeconomics back into the political economy of reform: the Lisbon Agenda and the 'fiscal philosophy' of EMU. JCMS 45 (1), 81–103.
Mahler, V.A., Jesuit, D.K., 2006. Fiscal redistribution in the developed countries: new insights from the Luxembourg Income Study. Soc. Econ. Rev. 4 (3), 483–511.
Mantovani, D., Papadopoulos, F., Sutherland, H., Tsakloglou, P., 2007. Pension incomes in the European Union: policy reform strategies in comparative perspective. In: Bargain, O. (Ed.), Micro-Simulation in Action: Policy Analysis in Europe using EUROMOD. Research in Labor Economics, vol. 25. Elsevier, Oxford, pp. 27–71.
Martinez-Vazquez, J., Rider, M., 2005. Multiple modes of tax evasion: theory and evidence. Natl. Tax J. 58 (1), 51–76.
Marx, I., Vandenbroucke, P., Verbist, G., 2012. Can higher employment levels bring down relative income poverty in the EU? Regression-based simulations of the Europe 2020 target. J. Eur. Soc. Policy 22 (5), 472–486.
Matsaganis, M., Flevotomou, M., 2008. A basic income for housing? Simulating a universal housing transfer in the Netherlands and Sweden. Basic Income Stud. 2 (2), 1–25.
Matsaganis, M., Leventi, C., 2013. The distributional impact of the Greek crisis in 2010. Fisc. Stud. 34 (1), 83–108.
Matsaganis, M., O'Donoghue, C., Levy, H., Coromaldi, M., Mercader-Prats, M., Rodrigues, C.F., Toso, S., Tsakloglou, P., 2006. Reforming family transfers in Southern Europe: is there a role for universal child benefits? Soc. Policy Soc. 5 (2), 189–197.
Matsaganis, M., Levy, H., Flevotomou, M., 2010. Non-take up of social benefits in Greece and Spain. Soc. Policy Adm. 44 (7), 827–844.
McFadden, D., 1974. Conditional logit analysis of qualitative choice behaviour. In: Zerembka, P. (Ed.), Frontiers in Econometrics. Academic Press, New York, pp. 105–142 (Chapter 4).

Merz, J., 1986. Structural adjustment in static and dynamic microsimulation models. In: Orcutt, G.H., Merz, J., Quinke, H. (Eds.), Microanalytic Simulation Models to Support Social and Financial Policy. North-Holland, Amsterdam, pp. 423–446.
Merz, J., 1991. Microsimulation—a survey of principles, developments and applications. Int. J. Forecast. 7 (1), 77–104.
Meyer, B.D., Sullivan, J.X., 2011. Further results on measuring the well-being of the poor using income and consumption. Can. J. Econ. 44 (1), 52–87.
Meyer, B.D., Mok, W.K.C., Sullivan, J.X., 2009. The under-reporting of transfers in household surveys: its nature and consequences, NBER Working Paper 15181.
Mirrlees, J., Adam, S., Besley, T., Blundell, R., Bond, S., Chote, R., Gammie, M., Johnson, P., Myles, G., Poterba, J. (Eds.), 2010. Dimensions of Tax Design: The Mirrlees Review. Oxford University Press, Oxford.
Mitton, L., Sutherland, H., Weeks, M. (Eds.), 2000. Microsimulation Modelling for Policy Analysis: Challenges and Innovations. Cambridge University Press, Cambridge.
Moffitt, R., 1983. An economic model of welfare stigma. Am. Econ. Rev. 73 (5), 1023–1035.
Morawski, L., Myck, M., 2010. 'Klin'-ing up: effects of Polish tax reforms on those in and on those out. Labour Econ. 17 (3), 556–566.
Navicke, J., Rastrigina, O., Sutherland, H., 2014. Nowcasting indicators of poverty risk in the European Union: a microsimulation approach. Soc. Indic. Res. 119 (1), 101–119.
Nelissen, J.H., 1998. Annual versus lifetime income redistribution by social security. J. Public Econ. 68 (2), 223–249.
Nolan, B., Callan, T., Maître, B., 2013. Country case study—Ireland. In: Jenkins, S.P., Brandolini, A., Micklewright, J., Nolan, B. (Eds.), The Great Recession and the Distribution of Household Income. Oxford University Press, Oxford, pp. 113–129 (Chapter 4).
OBR, 2013. Fiscal Sustainability Report. The Stationery Office, London.
O'Donoghue, C., 2001. Dynamic microsimulation: a methodological survey. Braz. Electron. Econ. J. 4, 1–77
O'Donoghue, C. (Ed.), 2014. Handbook of Microsimulation Modelling. Emerald, Bingley (forthcoming).
O'Donoghue, C., Baldini, M., Mantovani, D., 2004. Modelling the redistributive impact of indirect taxes in Europe: an application of EUROMOD, EUROMOD Working Paper EM7/01, University of Cambridge.
O'Donoghue, C., Lennon, J., Hynes, S., 2009. The Life-Cycle Income Analysis Model (LIAM): a study of a flexible dynamic microsimulation modelling computing framework. Int. J. Microsimulation 2 (1), 16–31.
OECD, 2007. Benefits and Wages 2007. OECD, Paris.
Orcutt, G.H., 1957. A new type of socio-economic system. Rev. Econ. Stat. 39 (2), 116–123.
Orcutt, G.H., Greenberger, M., Korbel, J., Rivlin, A., 1961. Microanalysis of Socio-Economic Systems: A Simulation Study. Harper and Row, New York.
Orcutt, G.H., Caldwell, S., Wertheimer, R., 1976. Policy Explorations Through Microanalytic Simulation. The Urban Institute, Washington, DC.
Paulus, A., Peichl, A., 2009. Effects of flat tax reforms in Western Europe. J. Policy Model 31 (5), 620–636.
Paulus, A., Čok, M., Figari, F., Hegedüs, P., Kralik, S., Kump, N., Lelkes, O., Levy, H., Lietz, C., Mantovani, D., Morawski, L., Sutherland, H., Szivos, P., Võrk, A., 2009. The effects of taxes and benefits on income distribution in the enlarged EU. In: Lelkes, O., Sutherland, H. (Eds.), Tax and Benefit Policies in the Enlarged Europe: Assessing the Impact with Microsimulation Models. Public Policy and Social Welfare, vol. 35. Ashgate, Vienna, pp. 65–90 (Chapter 4).
Paulus, A., Sutherland, H., Tsakloglou, P., 2010. The distributional impact of in-kind public benefits in European countries. J. Policy Anal. Manage. 29 (2), 243–266.
Pechman, J.A., 1973. Responsiveness of the federal individual income tax to changes in income. Brook. Pap. Econ. Act. 2, 385–427.
Peichl, A., 2009. The benefits and problems of linking micro and macro models—evidence from a flat tax analysis. J. Appl. Econ. 12 (2), 301–329.
Peichl, A., Siegloch, S., 2012. Accounting for labor demand effects in structural labor supply models. Labour Econ. 19 (1), 129–138.
Piketty, T., Saez, E., 2007. How progressive is the U.S. federal tax system? A historical and international perspective. J. Econ. Perspect. 21 (1), 3–24.

Pissarides, C.A., Weber, G., 1989. An expenditure-based estimate of Britain's black economy. J. Public Econ. 39, 17–32.
Popova, D., 2014. Distributional impacts of cash allowances for children: a microsimulation analysis for Russia and Europe. EUROMOD Working Paper EM2/14, University of Essex, Colchester.
Pudney, S., Sutherland, H., 1994. How reliable are microsimulation results? An analysis of the role of sampling error in a U.K. tax-benefit model. J. Public Econ. 53 (3), 327–365.
Pudney, S., Sutherland, H., 1996. Statistical reliability in microsimulation models with econometrically-estimated behavioural responses. In: Harding, A. (Ed.), Microsimulation and Public Policy. Contributions to Economic Analysis, vol. 232. North-Holland, Amsterdam, pp. 473–503 (Chapter 21).
Pudney, S., Hancock, R., Sutherland, H., 2006. Simulating the reform of means-tested benefits with endogenous take-up and claim costs. Oxf. Bull. Econ. Stat. 68 (2), 135–166.
Randelović, S., Rakić, J.Ž., 2013. Improving work incentives in Serbia: evaluation of a tax policy reform using SRMOD. Int. J. Microsimulation 6 (1), 157–176.
Redmond, G., Sutherland, H., Wilson, M., 1998. The Arithmetic of Tax and Social Security Reform. A User's Guide to Microsimulation Methods and Analysis. Cambridge University Press, Cambridge.
Riphahn, R.T., 2001. Rational poverty or poor rationality? The take-up of social assistance benefits. Rev. Income Wealth 47 (3), 379–398.
Robilliard, A.-S., Bourguignon, F., Robinson, S., 2008. Examining the social impact of the Indonesian financial crisis using a macro-micro model. In: Bourguignon, F., Bussolo, M., Pereira da Silva, L.A. (Eds.), The Impact of Macroeconomic Policies on Poverty and Income Distribution: Macro-Micro Evaluation Techniques and Tools. The World Bank and Palgrave Macmillan, New York, pp. 93–118 (Chapter 4).
Rodrigues, C.F., 2007. Income in EU-SILC—net/gross conversion techniques for building and using EU-SILC databases. In: Comparative EU, Statistics on Income and Living Conditions: Issues and Challenges. Eurostat Methodologies and Working Papers, Office for Official Publications of the European Communities, Luxembourg, pp. 157–172.
Salanauskaite, L., Verbist, G., 2013. Is the neighbour's grass greener? Comparing family support in Lithuania and four other New Member States. J. Eur. Soc. Policy 23 (3), 315–331.
Shorrocks, A., 2013. Decomposition procedures for distributional analysis: a unified framework based on the Shapley value. J. Econ. Inequal. 11 (1), 99–126.
Slemrod, J., 2007. Cheating ourselves: the economics of tax evasion. J. Econ. Perspect. 21 (1), 25–48.
Spielauer, M., 2011. What is social science microsimulation? Soc. Sci. Comput. Rev. 29 (1), 9–20.
Stiglitz, J.E., 2012. The Price of Inequality. W. W. Norton & Co., New York.
Sutherland, H., 1995. Static microsimulation models in Europe: a survey. Working Papers in Economics 9523, University of Cambridge.
Sutherland, H., 2014. Multi-country microsimulation. In: O'Donoghue, C. (Ed.), Handbook of Microsimulation Modelling. Emerald, Bingley (forthcoming).
Sutherland, H., Figari, F., 2013. EUROMOD: the European Union tax-benefit microsimulation model. Int. J. Microsimulation 6 (1), 4–26.
Sutherland, H., Taylor, R., Gomulka, J., 2002. Combining household income and expenditure data in policy simulations. Rev. Income Wealth 48 (4), 517–536.
Sutherland, H., Hancock, R., Hills, J., Zantomio, F., 2008. Keeping up or falling behind? The impact of benefit and tax uprating on incomes and poverty. Fisc. Stud. 29 (4), 467–498.
Tanton, R., Edwards, K.L. (Eds.), 2013. Spatial Microsimulation: A Reference Guide for Users. Springer, New York.
Tanton, R., McNamara, J., Harding, A., Morrison, T., 2009. Small area poverty estimates for Australia's Eastern Seaboard in 2006. In: Zaidi, A., Harding, A., Williamson, P. (Eds.), New Frontiers in Microsimulation Modelling. Ashgate, Vienna, pp. 79–95 (Chapter 3).
Tanton, R., Vidyattama, Y., Nepal, B., McNamara, J., 2011. Small area estimation using a reweighting algorithm. J. R. Stat. Soc. Ser. A Stat. Soc. 174 (4), 931–951.
Thoresen, T.O., 2004. Reduced tax progressivity in Norway in the nineties: the effect from tax changes. Int. Tax Public Financ. 11 (4), 487–506.

Urzúa, C.M. (Ed.), 2012. Fiscal Inclusive Development: Microsimulation Models for Latin America, Instituto Tecnológico y de Estudios Superiores de Monterrey (ITESM). International Development Research Centre, United Nations Development Programme.

van Soest, A., 1995. Structural models of family labor supply: a discrete choice approach. J. Hum. Resour. 30 (1), 63–88.

Verbist, G., 2007. The distribution effect of taxes on pensions and unemployment benefits in the EU-15. In: Bargain, O. (Ed.), Micro-Simulation in Action: Policy Analysis in Europe using EUROMOD. Research in Labor Economics, vol. 25. Elsevier, Oxford, pp. 73–99.

Verbist, G., Figari, F., 2014. The redistributive effect and progressivity of taxes revisited: an international comparison across the European Union. FinanzArchiv (forthcoming).

Waddell, P., Borning, A., Noth, M., Freier, N., Becke, M., Ulfarsson, G., 2003. Microsimulation of urban development and location choices: design and implementation of UrbanSim. Netw. Spat. Econ. 3 (1), 43–67.

Wagstaff, A., van Doorslaer, E., van der Burg, H., Calonge, S., Christiansen, T., Citoni, G., Gerdtham, U.-G., Gerfin, M., Gross, L., Häkinnen, U., John, J., Johnson, P., Klavus, J., Lachaud, C., Lauridsen, J., Leu, R.E., Nolan, B., Peran, E., Propper, C., Puffer, F., Rochaix, L., Rodríguez, M., Schellhorn, M., Sundberg, G., Winkelhake, O., 1999. Redistributive effect, progressivity and differential tax treatment: personal income taxes in twelve OECD countries. J. Public Econ. 72 (1), 73–98.

Wang, C., Caminada, K., Goudswaard, K., 2012. The redistributive effect of social transfer programmes and taxes: a decomposition across countries. Int. Soc. Secur. Rev. 65 (3), 27–48.

Wheaton, L., 2007. Underreporting of means-tested transfer programs in the CPS and SIPP. In: 2007 Proceedings of the American Statistical Association, Social Statistics Section. American Statistical Association, Alexandria, VA, pp. 3622–3629.

Wheaton, L., Giannarelli, L., Martinez-Schiferl, M., Zedlewski, S.R., 2011. How do States' safety net policies affect poverty? Working Families Paper 19, The Urban Institute, Washington, DC.

Whelan, S., 2010. The take-up of means-tested income support. Empir. Econ. 39 (3), 847–875.

Wilkinson, K., 2009. Adapting EUROMOD for use in a developing country—the case of South Africa and SAMOD. EUROMOD Working Paper EM5/09, University of Essex, Colchester.

Wolf, D.A., 2004. Book review of Microsimulation in Government Policy and Forecasting (2000). In: Gupta, A., Kapur, V. (Eds.), Journal of Artificial Societies and Social Simulation 7, (1).

Wolfson, M., 2009. Preface—Orcutt's vision 50 years on. In: Zaidi, A., Harding, A., Williamson, P. (Eds.), New Frontiers in Microsimulation Modelling. Ashgate, Vienna, pp. 21–29.

World Bank, 2013. Poverty prospects in Europe: assessing progress towards the Europe 2020 poverty and social exclusion targets in New European Union Member States. Report no: ACS4943, Human Development and Poverty Reduction and Economic Management Units.

Wright, G., Noble, M., Barnes, H., 2014. NAMOD: a Namibian tax-benefit microsimulation model. EUROMOD Working Paper EM7/14, University of Essex, Colchester.

Wu, B., Birkin, M., 2013. Moses: a dynamic spatial microsimulation model for demographic planning. In: Tanton, R., Edwards, K.L. (Eds.), Spatial Microsimulation: A Reference Guide for Users. Understanding Population Trends and Processes, vol. 6. Springer, New York, pp. 171–194.

Zaidi, A., Rake, K., 2001. Dynamic microsimulation models: a review and some lessons for SAGE. SAGE Discussion Paper 2, London School of Economics.

Zaidi, A., Harding, A., Williamson, P. (Eds.), 2009. New Frontiers in Microsimulation Modelling. Public Policy and Social Welfare, vol. 36. Ashgate, Vienna.

Zantomio, F., Pudney, S., Hancock, R., 2010. Estimating the impact of a policy reform on benefit take-up: the 2001 extension to the minimum income guarantee for UK pensioners. Economica 77 (306), 234–254.

INDEX

Note: Page numbers followed by *b* indicate boxes, *f* indicate figures and *t* indicate tables.

B

C

D

E

G

H

J

K

Q

R

S

Z